Patient Related Outcome Measures

Dovepress
open access to scientific and medical research

Open Access Full Text Article

REVIEW

Specific Measures of Quality of Life in Patients with Multimorbidity in Primary Healthcare: A Systematic Review on Patient-Reported Outcome Measures' Adequacy of Measurement

This article was published in the following Dove Press journal:
Patient Related Outcome Measures

Anne Møller
Kristine Henderson Bissenbakker
Anne Beiter Arreskov
John Brodersen

Research Unit for General Practice and Section of General Practice, Department of Public Health, University of Copenhagen, Copenhagen, Denmark

Purpose: The aim of this study is to search systematically for Patient Reported Outcome Measures (PROMs) used among patients with multimorbidity. Furthermore, the aim is to evaluate the adequacy and validity of the PROMs identified.

Design and setting: This systematic review follows the PRISMA guidelines. To assess the adequacy and validity of the identified PROMs the COSMIN Risk of Bias Checklist is used, more specifically a validation of the development, content validity, structural validity, and internal consistency of the PROMs.

Results: Four PROMs were identified in the primary search, and one was found from references. The sixth PROM was published after the primary search. None of the identified PROMs were aimed specifically at measuring the quality of life in patients with multimorbidity. According to the checklist, the development process and content validity were rated "adequate" in only one measure and "invalid"/"doubtful"/"inadequate" in the rest of the measures. The structural validity of the measures was rated "adequate" in four measures and "very good" in one. Regarding the internal consistency, two measures were rated doubtful and three "very good". None of the six PROMs reported analyses about invariant measurement. The COSMIN Risk of Bias Checklist proved easy to use; however, there are some concerns in the rating of bias, that are discussed further.

Conclusion: All six PROMs developed for patients with multimorbidity identified possessed inadequacy in their measurement properties. Therefore, the aim for the future is to develop a valid and adequate measure of the quality of life among patients with multimorbidity.

Keywords: comorbidity, burden of treatment, burden of disease, validity, psychometric properties

Introduction

Multimorbidity, often defined as two or more chronic diseases in the same patient, is known to affect patients' quality of life (QoL) negatively. Qualitative studies show that multimorbidity influences relationships, social-, and work life.[1,2] In quantitative research, patient-reported outcome measures, PROMs, are useful to assess the quality of life (QoL). QoL is normally assessed in different domains, eg, social, emotional, and physical domains.[3] In 2004, a systematic review examined the relationship between multimorbidity and QoL in a primary care setting and found that there is an inverse relationship between the individual's number of medical conditions and physical

Correspondence: Anne Møller Øster Farimagsgade 5, København DK-1353, Denmark
Tel +45 35327171
Email amoeller@sund.ku.dk

submit your manuscript | www.dovepress.com
DovePress
http://doi.org/10.2147/PROM.S226576

domains of QoL.[4] It could be hypothesized that this relationship was more pronounced and found in all domains (including social and psychological). However, the results could be explained by the PROMs' inadequacy of measuring QoL and not lack of relationship. In 2015, Smith and colleagues revised a systematic review[5] of intervention studies among patients with multimorbidity.[6] Only 18 studies were identified altogether. The authors concluded that there are still uncertainties about the effectiveness of interventions in this group, and it could be explained by the adequacy of the PROMs used. Outcome measures in these studies were besides PROMs, clinical or mental health outcomes, eg, blood tests and depression scores, utilization of health services, patient and provider behavior, treatment satisfaction, and economic costs. Eight studies used PROMs including a variety of generic measures of QoL. Generic measures of QoL are questionnaires such as the Euro-Qol Measures, eg, EQ-5D[7] and SF-36.[8] However, the generic questions may not be appropriate to patients with a high level of symptom burden, such as patients with multimorbidity. As an example, the EQ-5D is often used in general practice to measure QoL among patients with multimorbidity.[9–12] This can be problematic, as the EQ-5D includes a question concerning problems when doing "Usual activities (eg, work, study, housework, family or leisure activities)", as these activities might not be relevant to chronically ill patients with disabilities.[7]

Generic measures are often used to measure QoL, irrespective of the underlying diseases.[13] However, generic measures are not capable of taking variations among patients with different diagnoses groups, cultures or languages into account. This can lead to over- or underestimation, due to "differential item functioning" (DIF).[14] The risk of DIF is especially high among patients with multimorbidity, because it is a heterogeneous patient group, with many different combinations of diagnoses.[15]

When comparing patient perspectives in qualitative studies with the content of generic measures, there is a risk of bias since important aspects of QoL could be excluded. For assessment of potential effects of interventions among patients with multimorbidity in future research, we need PROMs with adequate measurement properties. Therefore, this study aimed to identify existing PROMs of QoL used in studies of patients with multimorbidity in primary healthcare, and to study the measurement adequacy of these measures, focusing on content validity, unidimensionality (structural validity) and invariant measurement of the PROMs.[14]

Materials and Methods

A protocol of the review was submitted to PROSPERO (Registration number: CRD42018090082 06.03.2018). The systematic review was conducted following the PRISMA guidelines.[16] The COSMIN (The COnsensus-based Standards for the selection of health Measurement INstruments) checklist and the COSMIN Risk of Bias checklist for systematic reviews of PROMs[17,18] were published during the process and used in the review. The checklist is a tool developed to assess the methodological quality of PROMs, and we used the checklist to assess the development process, content validity, structural validity (unidimensionality), and internal consistency.

Eligibility

We used three primary search terms (see Table 1 for definitions): Primary healthcare, quality of life, and multimorbidity. We did not want to specify or define QoL further as health-related quality of life, since the aim of the review was to include all potentially relevant measures. We defined patients with multimorbidity as patients diagnosed with two or more chronic diseases at the same time (but excluded studies including mental diseases only). We included articles where multimorbidity was mentioned as a term but not further defined, and we included papers indexed with "co-morbidity" since the term "multimorbidity" is a new index-word (introduced as MESH term in 2017, PubMed.gov). Co-morbidity is normally defined as the presence of one or more diseases in addition to a primary, "index"- disease. However, in epidemiological research, the distinction between complications, co-morbidity and multimorbidity is a challenge, and therefore all the terms were included.[19] No limitations were used in the primary search regarding language.

Information Sources

An information specialist assisted the primary search. We searched PubMed between 1st of January 1980 and 1st of March 2018, and we searched Psych Info and Embassy using the same search terms, searched on 1st of March 2018. See Appendix 1 for specific search strategy.

Process of Selection and Data Management

Stage 1: Two authors (AM and AA) performed the search individually and screened the titles of the primary search and selected relevant abstracts independently: all abstracts of relevance were read, and data about measures of QoL were extracted. Abstracts were registered in an excel-file summarizing title, authors, patient group, index-disease or definition

Table 1 Search Terms and Definitions

Search Field	Definition	Search Terms
Multimorbidity	The presence of two or more chronic diseases in one individual	"Comorbidity"[Mesh] Multimorbidities multimorbidity Multi morbidities multimorbidit* multi-morbidit* Comorbidit* comorbidity comorbidities
Primary care	The first level of contact in individuals with the health care system (Alma Ata declaration WHO 1979)	"Primary Health Care"[Mesh] "Family Practice"[Mesh] "General Practice"[Mesh] Primary health care Family medicine General practice Family practice
Quality of life	A generic concept reflecting concern with the modification and enhancement of life attributes, eg, physical, political, moral, social environment as well as health and disease (Definition in pub med)	Quality of Life"[Mesh] Quality of Life life quality patient reported outcome patient reported outcome measures "patientreported outcome measures" "patient-reported outcome measures" Patient Satisfaction"[Mesh] Patient Satisfaction "Patient Satisfaction"

of multimorbidity and outcome measures according to quality of life. If information was not available from the abstract, full versions of articles were obtained. Reference lists were searched for additional information.

Stage 2: The two authors, AM and AA, met and compared the results of stage 1. When discrepancies were found, each reference was studied, and the relevance was discussed until consensus was found.

Measures used for assessment of domains of QoL were registered. The specific measures regarding patients with multimorbidity in primary healthcare were studied thoroughly, and the validity was assessed.

Validity of PROMs

The COSMIN Risk of Bias checklist was used as a tool to assess the validity of the identified PROMs. The checklist includes ten boxes; however, in this study we focused on Box 1: PROM development (includes details on design, the use of conceptual models, existing questionnaires, literature review, and the use of cognitive interviewing and pilot tests), Box 2: Content validity (assessed by a critical examination of the structure of gathering data for the PROM following the checklist), Box 3: Structural validity (meaning to which degree the scores of a PROM is an adequate reflection of the construct measured), and Box 4: Internal consistency (how the items are related, is evaluated by following the questions in the checklist regarding statistical psychometric measures and analyses).[17] To assess the validity, a set of questions is included in each box. The lowest rating of any standard determines the overall rating ("the worse score counts"-principle).[18] Two authors (AM and KHB) performed the assessment of the validity following the checklist independently. Hereafter, the two authors compared their ratings and came to an agreement when differences in ratings were observed. In a few cases, consensus was not found, and the ratings were discussed with a third author JB.

Results

The search yielded 2421 hits and 156 of these were screened as potentially relevant by title. After reading all

abstracts, 80 full-text articles were left. Additional searches in Embase and Psych Info found 12 and 9 articles, respectively, but no additional PROMs were identified.

Specific measures: We identified four PROMs specifically developed to measure domains of QoL among patients with multimorbidity.[20–23] None of these PROMs have yet been evaluated using COSMIN guidelines according to the COSMIN database (searched 22.05.2018). A fifth PROM was identified from reference lists,[24] and a sixth was published after our main search[25] and appeared in our updated searches during work with this review. In the following, each of the six identified PROMs are evaluated according to the COSMIN checklist in relation to development, content validity, structural validity, and internal consistency.

Health Care Hassles Scale (HCHS)

The HCHS was developed among American war veterans with both single and multiple diseases in order to examine the relationship between attributes of primary care and healthcare system hassles.[22] The HCHS included questions regarding problems with seeking information and interacting with healthcare providers, problems with taking medications, and problems with accessing healthcare.

PROM Development and Content Validity

The origin of the construct to be measured and the conceptual framework were not described adequately. Furthermore, the target population of the designed PROM was unclear, both in the development process and for future assessments. In a pilot study, 26 items were tested in 132 primary care patients, but it was unclear whether the definition "primary care patients" covers "veterans with one or more chronic diseases", which was the population the PROM was aimed for. It was described how focus groups with patients with two or more chronic diseases were used to add further items to the item pool,[1] but the origin of the items was not described further. Therefore, we were not able to validate the content validity of the HCHS.

Structural Validity and Internal Consistency

The process of the mentioned "review of initial psychometric properties" before the final version of HCHS was not described. The final version included 16 items, and the items were coded on a 4-point scale and summarized in a total score of 0–64. Item-total correlations ranged from 0.45 to 0.81 with a Chronbach's alpha (α) at 0.94. Exploratory factor analysis (EFA) was used to examine structural validity.[22]

Measurement Adequacy According to COSMIN Checklist Box 1–4

The development process of the measure was not described adequately. The content validity of the HCHS was rated as "invalid", as it was not described. The analyses on structural validity and internal consistency were rated as "adequate", but since HCHS was based on an inadequate development process it was implicitly inadequate.

HealthCare Task Difficulty (HCTD)

Boyd et al aimed at developing a comprehensive measure of "HealthCare Task Difficulty" among elderly patients with multimorbidity.[23] HCTD covers themes about obtaining medication, planning medication schedules, administering medication, changing medications, problems with bills, scheduling appointments, transportation, and getting information. Furthermore, the measure includes questions about diet, medical equipment, and problems with community services.

PROM Development and Content Validity

The definition of multimorbidity was not specified. The HCTD was developed using 11 items about healthcare tasks generated from a literature review and input from clinical experts in geriatric medicine. Three items were excluded because they had an additional response category, compared with the 8 remaining items. The item development process was not described further: There was no information about whether items were tested separately on patients and professionals, nor about the methods and analyses used. There were references to a study about the feasibility of an intervention called: "Guided Care". However, this pilot study describes an intervention to enhance the quality of primary care experiences for chronically ill persons aged 65 and above. It did not describe the development of HCTD any further or tests of items.[26] Therefore, we were not able to assess HCTD's content validity.

Structural Validity and Internal Consistency

EFA was used to assess the dimensionalities of the HCTD, revealing one factor with an eigenvalue greater than 1. Afterwards, unidimensionality of this factor was verified by confirmatory factor analysis (CFA). α was 0.89 for all 8 items, although analyses found two subscales indicating multi-dimensionality. α was not calculated for these two potential subscales.

Measurement Adequacy According to COSMIN Box 1–4

The development and content validity of HCTD was rated as "invalid" due to lack of a development study/pilot study. The analyses on structural validity and internal consistency were rated as "adequate", but, when based on an inadequate development process, this was implicitly inadequate as well.

Patient-Reported Measure of Treatment Burden (PETS)

The authors of PETS aimed at assessing the burden of treatment among patients with multimorbidity in 11 content domains: Information, medication, appointments, monitoring, interpersonal challenges, expenses, difficulty with healthcare services, role/social activity, and physical/mental exhaustion.[20]

PROM Development and Content Validity

The questionnaire was developed from a conceptual model based on interviews with patients with multimorbidity.[27,28] Two authors developed 121 items based on the conceptual model. The items were reviewed by a stakeholder panel consisting of both clinicians and researchers,[20] and the draft was revised based on the feedback from the panel. A bank of 87 items was submitted to cognitive pretesting in two groups of relevant patients. After the first round of interviews, the item bank was revised, and the groups completed the revised items. There is no information on whether the interviewers were experienced, whether the interviews were transcribed, and how the analysis process was conducted. The final draft contained 78 items covering 15 content domains,[20] and it was validated in a survey among patients with multimorbidity.

Structural Validity and Internal Consistency

Based on classical test theory, items with missing data and items with content overlap and lack of conceptual fit were removed. Confirmatory factor analysis with the remaining 48 items in 9 content domains provided satisfactory fit with the data. Items were removed during the analysis. α values were 0.79–0.95 for the nine domains.

Measurement Adequacy According to COSMIN Checklist Box 1–4

The design process of PETS was rated as "very good", but because of lack of a few details on methods and analysis used in the cognitive interviews, the overall development process is rated as "adequate". There was no separate content validity study after the items were developed, but due to the thorough development of a conceptual framework including qualitative interviews and feedback from a multidisciplinary panel, we chose to rate the content validity as "adequate". The analysis on structural validity was rated as "adequate", and not "very good", because of too few participants included in relation to the number of items. The internal consistency was rated as "very good".

MULTimorbidity Illness Perceptions Scale (MULTIPleS)

MULTIPleS was developed specifically for patients with multimorbidity to assess illness perceptions. The aim of the measure was to assess the psychological processes underlying patient adjustment to multimorbidity. MULTIPleS included items about emotional representations, treatment burden, prioritization among conditions, causal links, and activity limitations.[21]

PROM Development and Content Validity

MULTIPleS was based on a model of illness perception and interviews with patients with multimorbidity (defined as patients with at least two out of five predefined chronic conditions) about their illness perceptions.[29] Based on an existing illness perception questionnaire and the results from qualitative work, 53 relevant items were identified. 11 items were removed after conducting cognitive interviews with 11 patients with multimorbidity. No further descriptions of the cognitive interviews or the interviewers' skills have been made. The remaining 42 items were conceptually grouped in five dimensions. It has not been mentioned whether stakeholders or specialists were included in the development process. There was no content validity study apart from assessment of face validity in the cognitive interviews.

Structural Validity and Internal Consistency

EFA was used for initial exploration of dimensionality. Then, Rasch analysis was used to evaluate the psychometric properties of the five sub-scales in MULTIPleS. During the analyses, items were removed and the Lickert scale scoring was changed. The participants were divided into two groups for, respectively, evaluation and validation. Finally, 5 separate scales (unidimensional scales) were combined to a summary scale with 22 items which showed "excellent fit with the Rasch model". No analyses on invariant measurement or DIF were reported despite the use of Rasch models. Moreover, analyses on α were 0.74–0.93 assessed for each scale.

Measurement Adequacy According to COSMIN Checklist Box 1–4

The development process is rated as "inadequate", since patients involved in the development process are not equivalent to the target group of the measure. The patients included in the qualitative interviews in the development process were patients with multimorbidity, their number being limited by their having at least two out of five predefined chronic conditions, whereas the target group of the PROM were patients without any mentioned limitations on included chronic conditions.

Furthermore, there was a lack of details in the development process including how the cognitive interviews were conducted and analyzed. As there was no separate content validity study of the items, apart from assessing face validity in the development process, this cannot be rated according to COSMIN. The analyses of structural validity and internal consistency were rated as "very good", but as MULTIPleS was based on an inadequate development process, its measurement adequacy was implicitly inadequate.

Treatment Burden Questionnaire (TBQ)

In 2012, Tran et al published a paper regarding an instrument to assess the burden of treatment among in- and outpatients with multiple chronic conditions and treatments,[24] including questions about drug intake, surveillance, lifestyle changes, and impact of healthcare on social relationships.

PROM Development and Content Validity

The TBQ was developed from items in existing questionnaires identified in a literature review and existing questionnaires and selected by members of the research team with experience in care of patients with chronic disease, but not specific stakeholders. No description of a conceptual model in the development process was made, but the underlying basis was the concepts of Burden of Treatment[28] and Minimal Disruptive Medicine.[30] Semi-structured interviews were performed with patients with at least one chronic disease from hospital or primary care. After the interviews, examples were added to the questions to increase comprehensibility, and the resulting questionnaire encompassed seven items, two of which had four sub-items. Finally, a panel of ten physicians (clinicians and methodologists) tested the face validity of the questionnaire and, apparently, no changes were made before tests of measurement properties.

Structural Validity and Internal Consistency

Item reduction was based on floor effect, the relevance of items and item redundancy by Spearman correlation coefficient. Due to results from these analyses, one sub-item was eliminated. The final questionnaire consisted of seven items (two of them with four sub-items) EFA was performed for examining dimensionality. It was described that factorial validity, assessed by scree plots, favored a unidimensional instrument, because 91% of the variance was explained by the first principal factor. α was 0.89.

Measurement Adequacy According to COSMIN Checklist Box 1–4

The development process was rated as "doubtful" due to lack of specifications of the qualitative process. Furthermore, the target population remained unclear; the aim was to develop a measure for treatment burden for patients with multiple conditions and treatments; however, patients involved in the development and validation process were stated as having at least one chronic condition (not necessarily all multiple conditions). The mean number of chronic conditions among the included patients was not reported, but according to the results, only 62.6% of patients had daily symptoms and some reported no main chronic condition. Inclusion of a variety of stakeholders for content validity was rated as adequate, but there was no separate content validity study among the patients. The test of structural validity was rated "adequate". The reduction of sub-items for the final questionnaire was not described adequately.

Multimorbidity Treatment Burden Questionnaire (MTBQ)

The aim of the MTBQ was to measure the burden of treatment among patients with multimorbidity, defined as patients having multiple long-term chronic conditions.[25] It was developed to assess the effect of interventions among patients with multimorbidity. The final scale included 10 items about medications, monitoring, lifestyle, dependence, and appointments with health professionals.

PROM Development and Content Validity

The PROMs concept of Burden of Treatment was well described, and the development of the scale included information from patients and a public involvement group. Furthermore, the authors included other measures of multimorbidity in the relevant patient group to get inspiration for items in the development process. The

scale was tested through a draft questionnaire including the HCTD questionnaire (mentioned above) in a pilot study with two rounds of cognitive interviews. The results from this pilot study were used for further testing of the items in a study called "the 3D study". The participants in the development process were eight patients above the age of 18 with a mean number of 2.1 diseases (1–5), whereas the questionnaire was validated among elderly patients with three or more chronic diseases (information found in).[25]

Structural Validity and Internal Consistency

EFA was performed to examine factors of the MTBQ and to conduct item reduction. These analyses led to questions about the relatedness of some items, but the authors chose to include them despite this, because of high content validity. The number of factors extracted was decided based on three criteria: 'eigenvalues greater than 1′, the scree plot, and by interpretability of domains not described any further. α was 0.83 in total.

Measurement Adequacy According to COSMIN Checklist Box 1–4

MTBQ development process was rated as "inadequate", due to the discrepancy between the target group of the measure: elderly patients with three or more chronic diseases and the patients used for the development. Furthermore, it was unclear which results were from the pilot test and from the final testing in the "3D study", respectively. The development process lacked details of methods in the interviewing process regarding transcription and information about whether the interviewers were skilled. These factors also influenced the COSMIN ratings of the development process.

Based on an "inadequate" development process, the structural validity and internal consistency of MTBQ were implicitly "inadequate". Furthermore, there was no further argumentation behind the decision of including items with doubtful relatedness, eg, reflection on results from the cognitive interviews.

Discussion

Six measures of patient-reported outcomes were identified, but none of them were aimed specifically at measuring the quality of life in patients with multimorbidity. The identified measures focused on difficulties in encounters with the healthcare system and in performing healthcare tasks, but also on burden of treatment and illness perception. It is obvious that these aspects influence a patients' quality of life, and therefore the measures were included in the review.

In this systematic review, we have used part of the COSMIN Risk of Bias checklist to assess the measurement adequacy of the identified PROMs. We found that the development process and content validity was rated "adequate" in only one of six measures (PETS) and invalid/doubtful/inadequate in the rest of the measures. The structural validity of the measures was rated "adequate" in HCHC, HCTD, PETS, TBQ, "very good" in MULTIPleS, and "inadequate" in MTBQ. According to the internal consistency, two measures were rated "doubtful" (MTBQ and HCHS) and three "very good" (HCTD, PETS, MULTIPLEs) (no available data from TBQ). Analyses of invariant measurement have not been performed in any of the six PROMs. This result will be elaborated on below.

The checklist proved easy to use as a relevant tool for systematically assessing the validity of the identified PROMs. However, in our opinion, the demands of fulfilling the checklist criteria are very high. Every item in the COSMIN checklist is rated equally, which has implications for the overall rating (see below). The concept of "worst score counts" also has a pronounced effect on the final grading of the validity of the PROM.

PROM Development and Content Validity

PROMs are used to evaluate health outcomes from the patient perspectives and should therefore be developed and tested in relevant patient groups. Thus, the COSMIN checklist regarding the development process includes questions about the inclusion of patients and stakeholders, as well as conceptual frameworks and models for the construct to be measured. In this systematic review, only three of the identified measures were tested among relevant patient groups (HCHS, PETS and MULTIPleS).[20–22]

In a review of PROMs of burden of treatment in patients with diabetic chronic kidney disease and heart failure, only 15 out of 57 identified measures included direct patient input.[31] In comparison, we have found a high degree of patient involvement in the development of new PROMs. In future development processes inclusion of professionals should also be prioritized.

None of the PROMs met the demands in the COSMIN checklist according to the development process. In our opinion, there are different explanations for that. One is the fact that each item in the checklist is rated equally. For instance, recording and transcription of the interviews is rated equally to description of the construct. Lack of

differentiation between items can lead to a low rating based on minor or less important deficits in the development process.

Structural Validity or unidimensionality

In the COSMIN checklist, methods used for test of unidimensionality defined as structural validity are: EFA, CFA, IRT and Rasch models. The authors of the COSMIN checklist value these psychometric models almost equally: CFA are valued "very good", while EFA are valued "adequate". If data fit IRT or Rasch models they are valued "very good". Most of the measures identified in this review used EFA as the only test of unidimensionality, even though the different types of EFA are all exploratory analyses and not confirmatory analyses, where one or more a priori hypotheses are tested. This makes EFA models very different from CFA, IRT and Rasch models, which are all confirmatory by nature. Therefore, we find it surprising that the COSMIN checklist weighs EFA and CFA equally in the validation of the structural validity. In addition, when conducting CFA, the data should also fit the model tested. Therefore, it is noteworthy that the COSMIN checklist only emphasizes model fit in IRT and Rasch models.

An example of the problems in using EFA for test of unidimensionality is seen in relation to the validation of MULTIPleS. There is a risk of losing potentially relevant items by using EFA to generate hypotheses, when there is a conceptual model as a base. In MULTIPleS, Rasch analyses are used afterwards to evaluate the psychometric properties of the five sub-scales identified. Moreover, the response categories in MULTIPleS are designed as a 6-point Likert scale from "strongly disagree" to "strongly agree". Disagreeing on a topic and agreeing on the same topic is not necessarily placed on unidimensional continuum measuring the same construct: it can be two or multi-dimensional.[32] This was not explored by Bower and colleagues.[21] Furthermore, disordered thresholds as revealed in Rasch analyses in the validation study of the MULTIPleS are not necessarily a problem with the response categories but can be signs of other anomalies, eg, multi-dimensionality, local response dependency, differential item function.[33] This was also not investigated by Gibbons et al.[21]

Rasch analyses are psychometric analyses based on the theory of item response theory (IRT). Compared to classic test theory, IRT is not based on an assumption of 1) normal distribution of data and 2) that responses to items are variables on an interval scale. Therefore, we believe IRT models should be rated superior to classic test theory models in the COSMIN checklist.[14]

Invariant Measurement and Internal Consistency

Rasch models are the only psychometric models that incorporate tests of invariant measurement in the analysis itself. Therefore, in our opinion, the EFA, CFA, IRT and Rasch models are not providing equal quality in measurement. The highest quality of measurement is provided with Rasch models, if items in a domain are shown to fit a Rasch model. In this case, the measure is shown to possess criterion-related construct validity,[34] to be objective,[35] sufficient,[36] and, therefore, also reliable.[37] Among the different IRT models, it is only the Rasch models that ensure sufficiency and invariant measurement ("specific objectivity" as George Rasch defined it).[35] If CFA and other IRT models are used to raise evidence of unidimensionality, additional analyses are needed for testing for invariant measurement. This is not mentioned or valued in the COSMIN Checklist.

Limitations

In the protocol, multimorbidity was not defined as including at least one physical disease. We found a couple of studies including patients with only psychiatric multimorbidity, which were not included in the final analysis, since we decided to include patients with at least one physical disease. One paper, Tran et al,[24] was not identified in our primary search, and therefore, hypothetically, we could have missed other relevant measures.

Since we have focused on the boxes in the checklist that were relevant for our purpose, it is possible that if we had evaluated the PROMs using all the boxes, this could have changed the evaluation.

Conclusion

We identified six patient-reported outcome measures developed for use among patients with multimorbidity. We found no measures specifically measuring the quality of life, but we did find PROMs assessing different aspects of healthcare problems, illness perception, and burden of treatment in the patient group. All six measures identified possessed inadequacy in their measurement properties according to the COSMIN Risk of Bias checklist. The checklist proved easy to use, but there are some concerns about the demands and priorities in the checklist.

Implication for Research

Due to the negative results of this systematic review, there is a need for development of a PROM assessing QoL among patients with multimorbidity. We will develop a new PROM for intervention studies among patients with two or more chronic diseases.

Based on a new conceptual model for QoL, existing PROMs, and interviews with patients with multimorbidity, we are now developing a PROM aiming for high content validity and adequate psychometric measurement properties with specific focus on unidimensional scales and invariant measurement among patients with multimorbidity.

Funding

AM is supported by a generous grant from the Novo Nordisk Foundation NNF15OC0019568.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Noël PH, Frueh BC, Larme AC, et al. Collaborative care needs and preferences of primary care patients with multimorbidity. *Health Expect*. 2005;8(1):54–63. doi:10.1111/hex.2005.8.issue-1
2. Ørtenblad L, Meillier L, Jønsson AR. Multi-morbidity: a patient perspective on navigating the health care system and everyday life. *Chronic Illn*. 2018;14(4):271–282. doi:10.1177/1742395317731607
3. Fayers PM, Machin D. *Quality of Life*. 3rd ed. West Sussex: Wiley; 2016.
4. Fortin M, Lapointe L, Hudon C, et al. Multimorbidity and quality of life in primary care: a systematic review. *Health Qual Life Outcomes*. 2004;2:51. doi:10.1186/1477-7525-2-51
5. Smith SM, Soubhi H, Fortin M, et al. Managing patients with multimorbidity: systematic review of interventions in primary care and community settings. *BMJ*. 2012;345:e5205. doi:10.1136/bmj.e5205
6. Smith SM, Wallace E, O'Dowd T, et al. Interventions for improving outcomes in patients with multimorbidity in primary care and community settings. *Cochrane Database Syst Rev*. 2016;3:CD006560.
7. EuroQol. EQ-5D. Available form: https://euroqol.org/. Accessed July 4, 2019.
8. Care RH. SF 36. Available form: https://www.rand.org/health/surveys_tools/mos/36-item-short-form.html. Accessed July 4, 2019.
9. Li J, Green M, Kearns B, et al. Patterns of multimorbidity and their association with health outcomes within Yorkshire, England: baseline results from the Yorkshire Health Study. *BMC Public Health*. 2016;16:649. doi:10.1186/s12889-016-3335-z
10. Mujica-Mota RE, Roberts M, Abel G, et al. Common patterns of morbidity and multi-morbidity and their impact on health-related quality of life: evidence from a national survey. *Qual Life Res*. 2015;24(4):909–918. doi:10.1007/s11136-014-0820-7
11. Vogel I, Miksch A, Goetz K, et al. The impact of perceived social support and sense of coherence on health-related quality of life in multimorbid primary care patients. *Chronic Illn*. 2012;8(4):296–307.
12. von dem Knesebeck O, Bickel H, Fuchs A, et al. Social inequalities in patient-reported outcomes among older multimorbid patients–results of the MultiCare cohort study. *Int J Equity Health*. 2015;14(17). doi:10.1186/s12939-015-0142-6.
13. Wacker ME, Jörres RA, Karch A, et al. Assessing health-related quality of life in COPD: comparing generic and disease-specific instruments with focus on comorbidities. *BMC Pulm Med*. 2016;16(1):70. doi:10.1186/s12890-016-0238-9
14. Brodersen J, Meads D, Kreiner S, et al. Methodological aspects of differential item functioning in the Rasch model. UK I, ed. *J Med Econ*. 2007;10:309–324. doi:10.3111/13696990701557048
15. Søndergaard E, Willadsen TG, Guassora AD, et al. Problems and challenges in relation to the treatment of patients with multimorbidity: general practitioners' views and attitudes. *Scand J Prim Health Care*. 2015;33(2):121–126. doi:10.3109/02813432.2015.1041828
16. Shamseer L, Moher D, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ*. 2015;349:g7647. doi:10.1136/bmj.g7647
17. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27:1147–1157. doi:10.1007/s11136-018-1798-3
18. Mokkink LB, de Vet HCW, Prinsen CAC, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2017.
19. Ording AG, Sørensen HT. Concepts of comorbidities, multiple morbidities, complications, and their clinical epidemiologic analogs. *Clin Epidemiol*. 2013;5:199–203. doi:10.2147/CLEP.S45305
20. Eton DT, Yost KJ, Lai J-S, et al. Development and validation of the Patient Experience with Treatment and Self-management (PETS): a patient-reported measure of treatment burden. *Qual Life Res*. 2016.
21. Gibbons CJ, Kenning C, Coventry PA, et al. Development of a multimorbidity illness perceptions scale (MULTIPleS). *PLoS One*. 2013;8(12):e81852. doi:10.1371/journal.pone.0081852
22. Parchman ML, Noël PH, Lee S. Primary care attributes, health care system hassles, and chronic illness. *Med Care*. 2005;43(11):1123–1129. doi:10.1097/01.mlr.0000182530.52979.29
23. Boyd CM, Wolff JL, Giovannetti E, et al. Healthcare task difficulty among older adults with multimorbidity. *Med Care*. 2014;52 Suppl 3:S118–S125. doi:10.1097/MLR.0b013e3182a977da
24. Tran V-T, Montori VM, Eton DT, et al. Development and description of measurement properties of an instrument to assess treatment burden among patients with multiple chronic conditions. *BMC Med*. 2012;10(1):68. doi:10.1186/1741-7015-10-68
25. Duncan P, Murphy M, Man M-S, et al. Development and validation of the Multimorbidity Treatment Burden Questionnaire (MTBQ). *BMJ Open*. 2018;8(4):e019413.
26. Boyd CM, Shadmi E, Conwell LJ, et al. A pilot test of the effect of guided care on the quality of primary care experiences for multimorbid older adults. *J Gen Intern Med*. 2008;23(5):536–542. doi:10.1007/s11606-008-0529-9
27. Eton DT, Ridgeway JL, Egginton JS, et al. Finalizing a measurement framework for the burden of treatment in complex patients with chronic conditions. *Patient Relat Outcome Meas*. 2015;6:117–126. doi:10.2147/PROM
28. Eton DT, Ramalho de Oliveira D, Egginton JS, et al. Building a measurement framework of burden of treatment in complex patients with chronic conditions: a qualitative study. *Patient Relat Outcome Meas*. 2012;3:39–49. doi:10.2147/PROM
29. Bower P, Harkness E, Macdonald W, et al. Illness representations in patients with multimorbid long-term conditions: qualitative study. *Psychol Health*. 2012;27(10):1211–1226. doi:10.1080/08870446.2012.662973
30. May C, Montori VM, Mair FS. We need minimally disruptive medicine. *BMJ*. 2009;339:b2803. doi:10.1136/bmj.b2803
31. Eton DT, Elraiyah TA, Yost KJ, et al. A systematic review of patient-reported measures of burden of treatment in three chronic diseases. *Patient Relat Outcome Meas*. 2013;4:7–20. doi:10.2147/PROM

32. Liu C-W, Chalmers RP. Fitting item response unfolding models to Likert-scale data using MIRT in R. *PLoS One*. 2018;13(5):e0196292. doi:10.1371/journal.pone.0196292
33. Andrich D. An expanded derivation of the threshold structure of the polytomous rasch model that dispels any "Threshold disorder controversy". *Educ Psychol Meas*. 2013;73(1):78–124. doi:10.1177/0013164412450877
34. Rosenbaum PR. Criterion-related construct validity. *Psychometrika*. 1989;54(4):625–633. doi:10.1007/BF02296400
35. Rasch G. An informal report on a theory of objectivity in comparisons. In: Van der Kamp LJTH, Vlek CAJ, editors. *Psychological Measurement Theory*. Leyden: University of Leyden; 1967.
36. Andersen EB. Sufficient statistics and latent trait models. *Psychometrika*. 1977;42(1):69–81. doi:10.1007/BF02293746
37. Bartholomew D. *The Statistical Approach to Social Measurement*. Academic Press. San Diego; 1996.