

REVIEW ARTICLE

# Scoping review of measures of treatment burden in patients with multimorbidity: advancements and current gaps

Daniel Mendoza-Quispe<sup>a,\*</sup>, Silvana Perez-Leon<sup>a</sup>, Christopher A. Alarcon-Ruiz<sup>b</sup>, Andrea Gaspar<sup>c</sup>, María Sofía Cuba-Fuentes<sup>d</sup>, Joseph R. Zunt<sup>e</sup>, Victor M. Montori<sup>f</sup>, Juan Carlos Bazo-Alvarez<sup>g</sup>, J. Jaime Miranda<sup>a,h,i</sup>

<sup>a</sup>CRONICAS Center of Excellence in Chronic Diseases, Universidad Peruana Cayetano Heredia, Lima, Peru

<sup>b</sup>Unidad de Investigación para la Generación y Síntesis de Evidencias en Salud, Universidad San Ignacio de Loyola, Lima, Peru

<sup>c</sup>School of Medicine, University of Washington, Washington, DC, USA

<sup>d</sup>Center for Research in Primary Health Care, Universidad Peruana Cayetano Heredia, Lima, Peru

<sup>e</sup>Departments of Neurology, Global Health, Medicine (Infectious Diseases), and Epidemiology, University of Washington, Seattle, WA, USA

<sup>f</sup>Knowledge and Evaluation Research Unit, Mayo Clinic, Rochester, MN 55905, USA

<sup>g</sup>Research Department of Primary Care and Population Health, University College London, London, UK

<sup>h</sup>School of Medicine, Universidad Peruana Cayetano Heredia, Lima, Peru

<sup>i</sup>The George Institute for Global Health, UNSW, Sydney, Australia

Accepted 16 May 2023; Published online 20 May 2023

## Abstract

**Objectives:** To identify, assess, and summarize the measures to assess burden of treatment in patients with multimorbidity (BoT-MMs) and their measurement properties.

**Study Design and Setting:** MEDLINE via PubMed was searched from inception until May 2021. Independent reviewers extracted data from studies in which BoT-MMs were developed, validated, or reported as used, including an assessment of their measurement properties (e.g., validity and reliability) using the Consensus-based Standards for the selection of health Measurement Instruments.

**Results:** Eight BoT-MMs were identified across 72 studies. Most studies were performed in English (68%), in high-income countries (90%), without noting urban-rural settings (90%). No BoT-MMs had both sufficient content validity and internal consistency; some measurement properties were either insufficient or uncertain (e.g., responsiveness). Other frequent limitations of BoT-MMs included absent recall time, presence of floor effects, and unclear rationale for categorizing and interpreting raw scores.

**Conclusion:** The evidence needed for use of extant BoT-MMs in patients with multimorbidity remains insufficiently developed, including that of suitability for their development, measurement properties, interpretability of scores, and use in low-resource settings. This review summarizes this evidence and identifies issues needing attention for using BoT-MMs in research and clinical practice. © 2023 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

**Keywords:** Chronic disease; Multimorbidity; Cost of illness; Surveys and questionnaires; Patient-reported outcome measures; Scoping review

Registration: The study protocol was registered in the Open Science Framework. DOI: <https://doi.org/10.17605/OSF.IO/FYKEJ>.

Availability of data and materials: All data supporting the conclusions of this manuscript are provided in the main tables, figures, and the supporting information.

Funding: This research is funded by the Fogarty International Center (FIC) at the US National Institutes of Health (GRANT 1R21TW011740-01). The funders did not play any role in relation to this work.

\* Corresponding author. CRONICAS Center of Excellence in Chronic Diseases, Universidad Peruana Cayetano Heredia, Av. Armendariz 445, Miraflores, Lima 18, Peru. Tel.: +5112416978.

E-mail address: [daniel.mendoza@upch.pe](mailto:daniel.mendoza@upch.pe) (D. Mendoza-Quispe).

## 1. Introduction

The prevalence of multimorbidity, defined as the occurrence of two or more chronic conditions, is rising throughout the globe, including low- and middle-income countries (LMICs) [1,2]. Patients with multimorbidity face multidimensional and often disruptive commitments to enact control over their multiple conditions. Each task loads additional time, resources, and effort, thus raising the burden attributed to their treatment [3]. A high burden of treatment (BoT) can lead to lower medication adherence and poorer health outcomes, in addition to feelings of isolation, loss of independence, stigma, and adverse physical effects [4].

**What is new?****Key findings**

- Treatment burden is the workload of care and the impact it has on the patient's quality of life. Multiple measures are available to assess burden of treatment in patients with multimorbidity (BoT-MMs), providing the basis for discriminating between low- and high-burdened patients.
- BoT-MMs were used mostly in English-speaking and high-income countries. There is practically no evidence on how these BoT-MMs are still valid in rural and low-resource settings.
- BoT-MMs have several technical limitations, such as suboptimal or under-investigated measurement properties (e.g., responsiveness), insufficient development (e.g., absent recall time, severe floor effect), and unclear rules for categorizing and interpreting raw scores (i.e., to clearly define high-burdened patients).

**What this adds to what was known?**

- This review identified eight BoT-MMs across 72 studies and, for the first time, summarizes their characteristics, their extent of use, and an assessment of their measurement properties with international standards.

**What is the implication and what should change now?**

- The evidence of BoT-MMS, summarized in this review, could guide clinicians, researchers, and future users to discern the most appropriate tool for their context and needs.

BoT is the workload experienced by patients due to the complexity of managing their medical care and its impacts on their functioning and well-being [5]. Its unique applicability relies on its potential to act as a patient-reported outcome of multimorbidity [6] and as a quality indicator of healthcare delivery that considers the patient's perspective, values, needs, and preferences [7]. This provides the basis for discriminating between low- and high-burdened patients [8], those who will experience the negative outcomes of being overwhelmed, and for measuring the impact of interventions on patients with a high burden.

Multiple measures, in the form of standardized questionnaires, were developed to assess BoT in patients with multimorbidity (BoT-MM) [9–11]; however, the evidence needed to inform the implementation of available BoT-MMs has been insufficiently examined. This evidence includes the characteristics of existing BoT-MMs and their

measurement properties, which express the ability of BoT-MMs to truly measure treatment burden (i.e., validity), to produce consistent results (i.e., reliability), and to capture changes over time (i.e., responsiveness). It also includes reviewing the context in which these BoT-MMs were previously developed, which is useful when planning studies on different contexts. For example, when a BoT-MM validated in a well-educated population from a high-income country is intended to be used in a low-income country setting with different education and access to healthcare services. Other reviews have focused on qualitative data [12], on multimorbidity quality of life [13], or on broad treatment burden, but not on patients with multimorbidity [14].

Thus, this review aimed to identify and assess available BoT-MMs, considering their measurement strengths and misuses. Our study has three objectives: 1) to describe the characteristics of the studies and populations in which available BoT-MMs were a) developed, reviewed, or adapted (e.g., in validation studies) or b) reported as used (i.e., in applicative studies); 2) to describe characteristics and measurement properties of available BoT-MMs, assessing the evidence from validation studies using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) [15,16]; and 3) to describe BoT estimates and other measurement products of BoT-MMs from applicative studies.

**2. Methods**

A scoping review was conducted under the methodological framework proposed by Arksey H. and O'Malley L [17] and the Preferred Reporting Items for Systematic reviews and Meta-Analyses- Scoping review (ScR) guidelines (see checklist on Table A.1) [18]. The study protocol was published elsewhere [19].

*2.1. Search strategy*

First review: MEDLINE via PubMed was searched from inception until May 2021 (see search formula in Table A.2). No language, year, or publication type restrictions were considered.

Post hoc review: PubMed was again searched to retrieve studies that cited any validation study included during the previous review, as we hypothesized that studies using BoT-MMs would cite validation studies. In addition, hand-searching involved screening the citations of included studies.

*2.2. Study selection*

We included studies in which BoT-MMs were a) developed, validated, or adapted (validation studies) or b) reported as used (applicative studies). We defined multimorbidity as having  $\geq 2$  chronic conditions. In

addition, we included instruments that were developed in patients with  $\geq 1$  chronic condition but were not disease-specific and thus could be applicable to individuals with multimorbidity. Studies with disease-specific questionnaires or nonstandard measures of treatment burden, protocols, systematic reviews, and case reports were excluded.

Records from literature searches were uploaded to Rayyan's online software (<https://rayyan.ai/>). Two reviewers (D.M-Q. and S.P-L.) independently performed the study selection and data extraction; discrepancies were resolved by consensus between reviewers. First, titles and abstracts were screened; then, articles were full-text reviewed to determine their inclusion. Previously, three calibrations were conducted to refine the application of eligibility criteria.

### 2.3. Data extraction

A charting data form in Microsoft Excel was designed by two reviewers (D.M-Q. and S.P-L.) and refined by two professionals with a background in validation studies. Data extracted included the following:

- (i) Characteristics of studies: general characteristics (title, first author, year of publication), study aim (validation vs. applicative-only studies), language, country (grouped by world regions and by income), place of residence (urban, rural), setting type (community, primary, secondary, and tertiary care, or mixed if  $\geq 1$  type setting), data on assistance during BoT-MM administration (self-reported, assisted, or both), and mode of administration (in person, remote [by post, telephone, and online], or both).
- (ii) BoT-MM's characteristics: language availability (original and translations), domains, constructs measured (treatment burden, a domain of it, or another construct), number of items, scores by item, range of scores (minimum-maximum), standards for interpreting raw scores (e.g., test norms from population-based reference scores), recall time, and reporting of ceiling and floor effects. Measurement properties included content validity, structural validity, internal consistency, reliability, construct validity, known-group validity, and responsiveness, according to COSMIN standards [16]. Data was collected for the first version and subsequent versions of BoT-MMs.
- (iii) BoT estimates and study samples: numerical (total and by subscales) or categorical (%) scores, and study sample characteristics including sample size, age, female sex (%), target conditions, education level, and ethnicity.

### 2.4. Assessment of measurement properties and COSMIN standards

Evidence from validation studies was reviewed independently by two reviewers (D.M-Q. and C.A.A-R.) using the

COSMIN methodology [15,16], in which measurement properties and their quality of evidence, that is, confidence that methods of each study ensure the accuracy of measurement properties reported, were summarized and rated according to standard criteria (Tables A.3 and A.4). A reflective measurement model, as defined by COSMIN, was assumed for all instruments. Briefly, the methodological quality of single studies was evaluated using the COSMIN Risk of Bias checklist; then, measurement properties per BoT-MM in each single study were collected (single result) and rated (single rating) as sufficient (+), insufficient (−), inconsistent ( $\pm$ ), or indeterminate (?) according to COSMIN standards (Tables A.5 and A.6). Measurement properties per BoT-MM were qualitatively summarized (overall result) and rated (overall rating), and their quality of evidence was graded as high, moderate, low, or very low using a modified Grading of Recommendations, Assessment, Development, and Evaluations approach (Table A.7).

### 2.5. Data synthesis and analysis

First, the study selection was summarized with a Preferred Reporting Items for Systematic reviews and Meta-Analyses flowchart. The interrater reliability between reviewers was assessed with Cohen's kappa. The characteristics of included studies were described by study aim (validation vs. applicative). A world map and a line chart depicting the number of studies by country and the number of studies per year, respectively, were generated in the program 'R' (ggplot2 library) to describe worldwide research and publication trends over time, respectively.

Then, characteristics and measurement properties per BoT-MMs were summarized. Each measurement property was summarized with overall results, overall ratings (+, −,  $\pm$ , or ?), and their quality of evidence (high, moderate, low, very low).

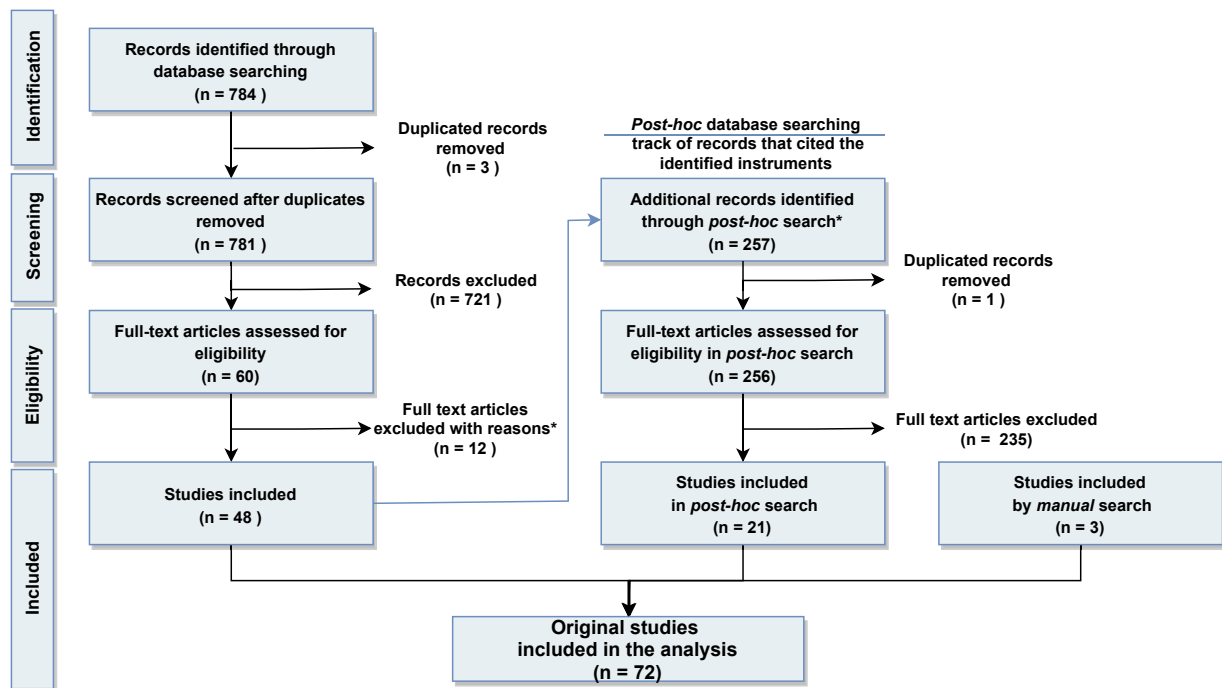
Lastly, summary scores obtained by single applicative studies were organized per BoT-MM. Numerical scores were summarized with mean (standard deviation [SD]) or median (interquartile range or range), as available, and categorical scores were reported according to cutoffs from Table A.8. If only subgroup scores were available, the aggregated mean  $\pm$  SD (or median) was estimated.

## 3. Results

A total of 72 studies linked to 8 BoT-MMs were included (Fig. 1). Overall, reviewers' agreement during full-text review was 94.9% (Cohen's kappa 0.86; Table A.9). Full data extraction is available in Appendix B.

### 3.1. Characteristics of included studies ( $n = 72$ )

Table 1 summarizes included studies; these were published from 2012 onwards, mostly (73%) between 2018



**Fig. 1.** Study flowchart. \*Excluded because the patient-reported outcome measure of treatment burden was i) disease-specific ( $n = 5$ ), ii) nonstandardized ( $n = 2$ ), or iii) not applicable ( $n = 4$ ). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

and 2021, depicting an increasing trend over time (Fig. 2). One-third (33%) were validation studies, and two-thirds (67%) were applicative studies. Most studies were conducted in Europe or North America (75%) using surveys in English language (68%). The predominant location of these studies was high-income countries (90%), with few conducted in middle-income countries (10%) and none in low-income countries. Place of residence was often not reported (90%), while the most common setting was primary care (46%). BoT-MMs were administered in person (43%) or remotely (41%), usually self-administered (67%), but in a third of studies (30%), the administration was assisted. Highly educated samples of white ethnicity predominated.

### 3.2. Characteristics of BoT-MMs tools ( $n = 8$ )

Table 2 provides an overview and full names of available BoT-MMs: Treatment Burden Questionnaire (TBQ) [9], Health Care Task Difficulty Questionnaire (HCTD) [43], Multimorbidity Illness Perceptions Scale (MULTIPLES) [35], Living with Medicines Questionnaire (LMQ) [37], National Health and Aging Trends Study (NHATS) [44], Patient Experience with Treatment and Self-management (PETS) [11], Medication-Related Burden Quality of Life (MRB-QoL) [36], and the Multimorbidity Treatment Burden Questionnaire (MTBQ) [10]. All BoT-MMs (100%) were available in English [10,11,20,28,34–36,38,39], and some BoT-MMs had multiple versions of varying length or language (TBQ [9,20–23], PETS [11,28–31], LMQ) [37–41]. Four BoT-

MMs (TBQ, MTBQ, PETS, NHATS) [9–11,44] assess treatment burden in full; another three (HCTD, MRB-QoL, LMQ) [34,36] assess a specific domain of treatment burden; and one BoT-MM (MULTIPLES) [35] assesses another construct with an embedded BoT subscale. About scores, higher scores reflected a higher BoT in all BoT-MMs, and scores were available as total scores (50%), subscale scores (50%), or categorical scores (50%). The rationale behind categorization of scores was often unclear (Table A.8). Only two (25%) BoT-MMs (PETS [11,28,32] and MRB-QoL [36]) included recall times. Six (75%) BoT-MMs had evidence of floor [9–11,20,22,26–28,35,36,39] or ceiling [36,38,39] effects.

### 3.3. Measurement properties of BoT-MMs ( $n = 7$ )

These properties are summarized in Table 3 (see the full item by item COSMIN assessment in Appendix C). One BoT-MM (NHATS) was not included in this assessment due to a lack of validation studies.

Four of nine measurement properties (content validity, internal consistency, structural validity, and construct validity) were investigated in all BoT-MMs ( $n = 7$ ). Content validity was sufficient in four BoT-MMs (MTBQ, PETS, MRB-QoL, LMQ) and insufficient in three BoT-MMs due to insufficient comprehensibility (TBQ) [9,20], or comprehensiveness (HCTD [34], MULTIPLES) [35], referred to asking patients for their understanding of instructions, items and answers given the wording of instruments or if they feel that all key aspects of BoT are covered by items,

**Table 1.** Summary of included studies ( $n = 72$ )

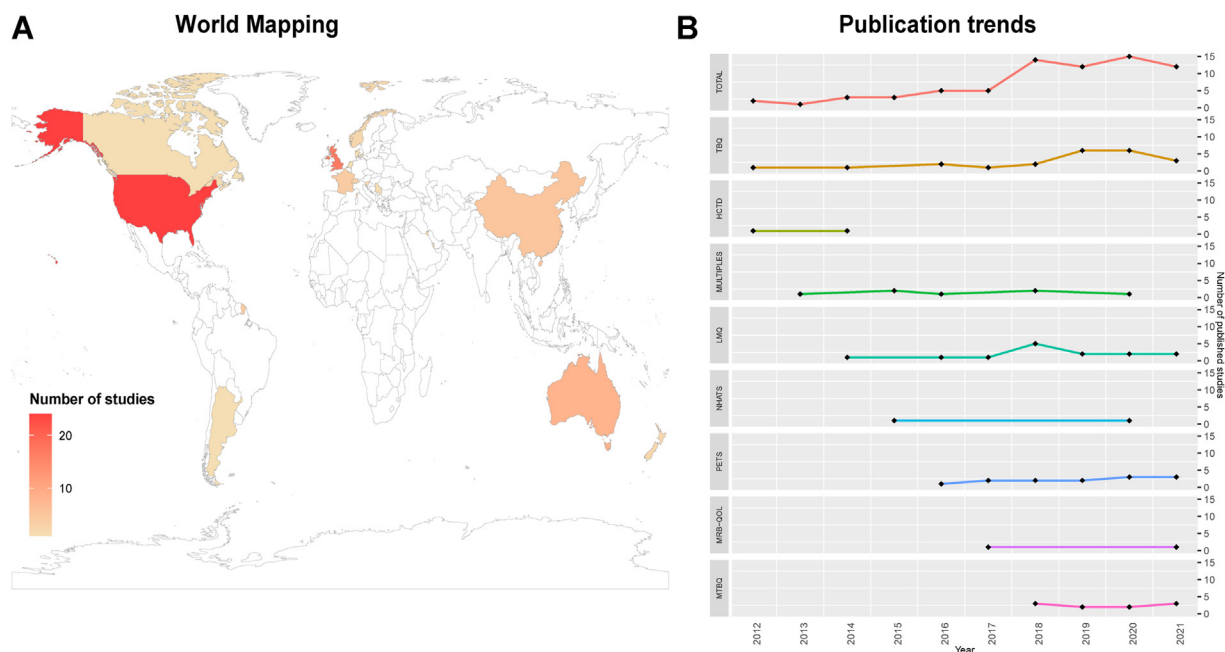
Characteristic	Validation		Applicative only		Total	
	$n = 24$	%	$n = 48$	%	$n = 72$	%
Region <sup>a</sup>						
Europe	8	34.8%	22	46.8%	30	42.9%
North America	7	30.4%	16	34.0%	23	32.9%
Oceania	2	8.7%	6	12.8%	8	11.4%
Asia	4	17.4%	3	6.4%	7	10.0%
South America	1	4.3%	-	-	1	1.4%
Multiple	1	4.3%	-	-	1	1.4%
Country income <sup>a</sup>						
High income	18	78.3%	45	95.7%	63	90.0%
Middle income	5	21.7%	2	4.3%	7	10.0%
Low income	-	-	-	-	-	-
Language <sup>a</sup>						
English	14	58.3%	32	72.7%	46	67.6%
Multiple	0	0.0%	5	11.4%	5	7.4%
Chinese	4	16.7%	1	2.3%	5	7.4%
French	2	8.3%	2	4.5%	4	5.9%
Others	4	16.7%	4	9.1%	8	11.8%
Place of residence						
Not reported	19	79.2%	41	93.2%	60	88.2%
Only urban	3	12.5%	3	6.8%	6	8.8%
Both reported	3	12.5%	3	6.8%	6	8.8%
Only rural	-	-	-	-	-	-
Ethnicity						
Reported	11	45.8%	30	62.5%	41	56.9%
Not reported	13	54.2%	18	37.5%	31	43.1%
Setting <sup>a</sup>						
Primary care	25	53.2%	7	30.4%	32	45.7%
Secondary or tertiary care	6	12.8%	6	26.1%	12	17.1%
Community	8	17.0%	5	21.7%	13	18.6%
Mixed	8	17.0%	5	21.7%	13	18.6%
Place of administration <sup>a</sup>						
In person	7	31.8%	22	47.8%	29	42.6%
Remote (online, post, phone)	10	45.5%	18	39.1%	28	41.2%
Both	5	22.7%	6	13.0%	11	16.2%
Mode of administration <sup>a</sup>						
Self-administered	14	66.7%	31	67.4%	45	67.2%
Assisted	5	23.8%	15	32.6%	20	29.9%
Both	2	9.5%	-	-	2	3.0%

<sup>a</sup> Totals may not add 72 due to missing data.

respectively. Internal consistency was sufficient in two BoT-MMs (HCTD, MULTIPLES), but indeterminate in five BoT-MMs (TBQ, MTBQ, MRB-QoL, PETS [11,28,30–32], and the LMQ) [38–40] because sufficient structural validity was not met despite internal consistency indicators (Cronbach alpha) being reported. Structural validity was sufficient in two BoT-MMs (HCTD, MULTIPLES) and insufficient in other BoT-MMs due to either inconsistent factor structures (TBQ: one- [9,20] or three-

factors [21–23]; MTBQ: one- [10] or three-factors [27]; PETS: nine- [11] or twelve-factors) [32], or indicators of factorial analysis below standard criteria (LMQ) [38–40] or not reported (MRB-QoL) [36]. Construct validity was sufficient in most BoT-MMs (except the PETS, which had sufficient ratings in only 4 of 12 subscales), which occurred in cross-sectional studies either when BoT-MMs scores correlated well with expected scores of reference measures (e.g., higher BoT-MMs scores were associated





**Fig. 2.** Mapping of research on measures of treatment burden in patients with multimorbidity. The number of studies are shown by (A) country and (B) over time. TBQ, Treatment Burden Questionnaire; MTBQ, Multimorbidity Treatment Burden Questionnaire; PETS, Patient Experience with Treatment and Self-management; HCTD, Health Care Task Difficulty Questionnaire; MULTIPLES, Multimorbidity Illness Perceptions Scale; NHATS, National Health and Aging Trends Study; MRB-QoL, Medication-Related Burden Quality of Life; LMQ, Living with Medicines Questionnaire. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

with lower quality of life) [10,20,39] or when known subgroups had significantly different scores (e.g., patients with low literacy had higher BoT-MM scores compared to those with adequate literacy) [11].

The remaining five of nine measurement properties were investigated in five (reliability), three (responsiveness), or less/none (measurement error, criterion validity, and cross-cultural validity/measurement invariance) of seven BoT-MMs. Reliability was sufficient in three BoT-MMs (TBQ, MTBQ, and LMQ) and most (9 of 12) PETS subscales [32], and insufficient in the MULTIPLES [35] due to overall results below standard criteria. Responsiveness was sufficient in one BoT-MMs (MTBQ) [10], unclear in most PETS subscales [30,31], and insufficient in one BoT-MM (HCTD) [34]. It was evaluated in longitudinal studies of 6 to 18 months of follow-up by comparing whether changes in treatment burden scores over time (follow-up baseline) were in accordance with expected changes in measures of reference variables, for example, quality of life [10,30,31,34].

### 3.4. BoT estimates and study samples ( $n = 58$ )

Applicative studies showed that summary scores, shown in separate tables per BoT-MM (Table 4 and Tables A.10–A.15), are available mostly for adults with various conditions, including  $\geq 1$  or  $\geq 2$  chronic conditions or single specific conditions [21,26,28,33,48,51,53,55–63]. The referred tables also show the range of mean population scores per BoT-MM; for example, mean TBQ scores

between studies ranged between 20.9 and 56.2 (scale range: 0–150). There was no equivalence of scores between BoT-MMs.

## 4. Discussion

This review summarizes the available evidence of eight BoT-MMs identified across 24 validation studies and 48 only-applicative studies. Our results showed that research was limited in low-resource settings (LMIC, urban-rural disparities), and extant BoT-MMs have common limitations such as several suboptimal or under-investigated measurement properties, insufficient development (absent recall time, presence of floor effects), and unclear rationale for categorizing and interpreting raw scores. We summarize this evidence and identify issues and gaps needing attention for using BoT-MMs in research and practice.

Our study revealed a growing interest in using BoT-MMs. However, most evidence arises from high-income countries, for example, USA and UK, whereas LMIC settings are underrepresented, including Latin America, Africa, Western Europe, and parts of Asia. Noteworthy, BoT measurement is strongly influenced by the surrounding health system and cultural appropriation of patients in each country, an important reflection given that most BoT-MMs were developed in highly educated populations with access to long-term healthcare, contrary to LMIC settings, which have unguaranteed continuity of care [6], major socioeconomic disparities, limited health literacy, and higher

**Table 2.** Overview of measures of treatment burden in patients with multimorbidity (BoT-MMs) ( $n = 8$ )

BoT-MM (reference)	Population	Original language	Translation	Main construct measured	Factor structure	Validated versions <sup>a</sup> (n)	Items (n)	Score by item	Range of scores (min - max)	Type of t scores	Ceiling or floor effect	Recall time	Number of studies (n)
TBQ [9,20–25]	MM, CC, MS [21], Stroke [23]	French	English, Spanish, Chinese	TB	One or three	e-TBQ-13 e-TBQ-15 s-TBQ-16 c-TBQ-15 mc-TBQ-15	13–16	0–10	0–130 0–150 0–160	Total, categorical	Floor [9,20, 22,26]	-	22
MTBQ [10,27]	MM	English	Chinese	TB	One	MTBQ-10 c-MTBQ-10	10 (+3)	0–4	0–100 (mean score*25)	Total, categorical	Floor [10,27]	-	10
PETS [11,28–33]	MM, CC Diabetes [28], HF [29], Cancer [29]	English	Norwegian	TB	Nine or twelve	e-PETS-60 n-PETS-59 e-PETS-48 e-PETS-34	34, 48, 59 or 60	0–5	0–100 (standardized)	By subscale	Floor [11,28]	4 wk	13
HCTD [34]	MM	English	-	A dimension of TB	One	e-HCTD-8	8 (+3)	0–2	0–16	Categorical	-	-	2
MULTIPLES [35]	MM	English	-	Illness perceptions	Five & unidimensional	e-MULTIPLES-22	22	0–5	0–78	Total, by subscale	Floor [35]	-	7
NHATS	MM	English	-	TB	?	e-NHATS-4	4	0–4	-	Categorical	-	-	2
MRB-QoL [36]	CC & regular medication	English	-	A dimension of TB	Five	e-MRB-QoL-31	31	0–5	0–100 (standardized)	By subscale	Floor & ceiling [36]	2 wk	2
LMQ [37–42]	CC & regular medication	English	Arabic, Chinese, Slovenian	A dimension of TB	Eight	e-LMQ-60 e-LMQ-42 e-LMQ-41 a-LMQ-41 c-LMQ-39 sl-LMQ-41	39, 41, 42 or 60	1–5	39–195 41–205 42–210 60–300	Total, by subscale, categorical	Ceiling [38]; Floor & ceiling [39]	-	14

*Abbreviations:* a, Arabic; e, English; c, Chinese; mc, Mandarin Chinese; n, Norwegian; s, Spanish; sl, Slovenian; TBQ, Treatment Burden Questionnaire; MTBQ, Multimorbidity Treatment Burden Questionnaire; PETS, Patient Experience with Treatment and Self-management; HCTD, Health Care Task Difficulty questionnaire; MULTIPLES, Multimorbidity Illness Perceptions Scale; MRB-QoL, Medication-Related Burden Quality of Life; LMQ, Living with Medicines Questionnaire; BoT-MMS, measures to assess burden of treatment in patients with multimorbidity; CC, chronic condition (s); HF, heart failure; MM, multimorbidity; MS, multiple sclerosis; TB, treatment burden.

<sup>a</sup> Notation: first initial of language-measure-number of items (e.g., e-TBQ-13: TBQ validated in English, with 13 items).

**Table 3.** Measurement properties by BoT-MM ( $n = 7$ ), COSMIN Standards

Measure	Content validity			Internal consistency	Structural validity	Reliability	Measurement error	Criterion validity	Construct validity	Responsiveness
	Relevance	Comprehensiveness	Comprehensibility							
	Rating (QoE)	Rating (QoE)	Rating (QoE)							
TBQ (total scale)	± (VL)	+ (M)	− (VL)	? (NA) <sup>a</sup>	± (L)	+ (VL)	? (NA)	? (NA)	+ (L)	
MTBQ (total scale)	+ (VL)	+ (L)	± (VL)	? (NA) <sup>a</sup>	? (NA)	+ (VL)		? (NA)	+ (M)	+ (M)
PETS (total scale)					± (VL)					
S1: Medical information	+ (VL)	+ (M)	+ (L)	? (NA) <sup>a</sup>		+ (VL)			+ (H)	
S2: Medications	+ (VL)	+ (M)	+ (L)	? (NA) <sup>a</sup>		− (VL)			+ (H)	
S3: Medical appointments	+ (VL)	+ (M)	+ (L)	? (NA) <sup>a</sup>		+ (VL)			− (M)	
S4: Monitoring health	+ (VL)	+ (M)	+ (L)	? (NA) <sup>a</sup>		+ (VL)			− (M)	
S5: Interpersonal challenges	+ (VL)	+ (M)	+ (L)	? (NA) <sup>a</sup>		− (VL)			− (M)	
S6: Medical expenses	+ (VL)	+ (M)	+ (L)	? (NA) <sup>a</sup>		+ (VL)			+ (M)	− (L)
S7: Difficulty with healthcare services	+ (VL)	+ (M)	+ (L)	? (NA) <sup>a</sup>		+ (VL)			− (L)	− (M)
S8: Social activity limitations	+ (VL)	+ (M)	+ (L)	? (NA) <sup>a</sup>		+ (VL)			− (M)	+ (M)
S9: Physical/mental exhaustion	+ (VL)	+ (M)	+ (L)	? (NA) <sup>a</sup>		+ (VL)			+ (H)	
S10: Difficulties performing exercise	+ (VL)	+ (M)	+ (L)	? (NA) <sup>a</sup>		+ (VL)			− (L)	− (M)
S11: Difficulties performing diets	+ (VL)	+ (M)	+ (L)	? (NA) <sup>a</sup>		+ (VL)			− (M)	+ (M)
S12: Medical equipment	+ (VL)	+ (M)	+ (L)	? (NA) <sup>a</sup>		− (VL)				
HCTD (total scale)	± (VL)	− (VL)	? (NA)	+ (M)	+ (M)				+ (L)	− (VL)
MULTPLES (total scale)	± (VL)	− (VL)	+ (L)	+ (M)	+ (M)	− (VL)			+ (M)	
S1: Treatment burden	± (VL)	− (VL)	+ (L)	+ (M)		− (VL)			+ (M)	
MRB-QoL (total scale)					? (NA)					



Table 3. Continued

Measure	Content validity			Internal consistency	Structural validity	Reliability	Measurement error	Criterion validity	Construct validity	Responsiveness
	Relevance	Comprehensiveness	Comprehensibility							
	Rating (QoE)	Rating (QoE)	Rating (QoE)							
S1: Routine and regimen complexity	+ (VL)	+ (VL)	? (NA)	? (NA) <sup>a</sup>					+ (H)	
S2: Psychological burden	+ (VL)	+ (VL)	? (NA)	? (NA) <sup>a</sup>					+ (H)	
S3: Functional and role limitation	+ (VL)	+ (VL)	? (NA)	? (NA) <sup>a</sup>					+ (H)	
S4: Therapeutic relationship	+ (VL)	+ (VL)	? (NA)	? (NA) <sup>a</sup>					+ (H)	
S5: Social burden	+ (VL)	+ (VL)	? (NA)	? (NA) <sup>a</sup>					– (H)	
LMQ (total scale)					– (L)					
S1: Relationships with healthcare professionals about medicines	± (VL)	+ (VL)	+ (VL)	? (NA) <sup>a</sup>		+ (VL)			+ (M)	
S2: Practical difficulties	± (VL)	+ (VL)	+ (VL)	? (NA) <sup>a</sup>		+ (VL)			+ (M)	
S3: Lack of effectiveness of medicines	± (VL)	+ (VL)	+ (VL)	? (NA) <sup>a</sup>		+ (VL)			+ (M)	
S4: Cost-related burden	± (VL)	+ (VL)	+ (VL)	? (NA) <sup>a</sup>		+ (VL)			+ (M)	
S5: Side effects of prescribed medications	± (VL)	+ (VL)	+ (VL)	? (NA) <sup>a</sup>		+ (VL)			+ (M)	
S6: Attitudes/concerns about medicine use	± (VL)	+ (VL)	+ (VL)	? (NA) <sup>a</sup>		+ (VL)			+ (M)	
S7: Impact/interference to day-to-day life	± (VL)	+ (VL)	+ (VL)	? (NA) <sup>a</sup>		+ (VL)			+ (M)	
S8: Control/autonomy to vary regimen).	± (VL)	+ (VL)	+ (VL)	? (NA) <sup>a</sup>		+ (VL)			– (M)	

Abbreviations: (+), Sufficient; (–), insufficient; (±), inconsistent; (?), indeterminate results; S, subscale; QoE, quality of evidence; H, high; M, moderate; VL, very low; L, low; NA, not applicable.

Gray background is shown if results were “–”, “±”, “?”, or there was no data (and data was expected).

<sup>a</sup> Internal consistency indicators (Cronbach alpha) were reported but rated as indeterminate (“?”) because sufficient structural validity was not met.

**Table 4.** Summary scores in studies using the Treatment Burden Questionnaire (TBQ)

First author, year of publication (reference)	Country	Condition	Setting	Sample size (n)	Age mean (SD)	Women (%)	Range of scores <sup>b</sup> (min/max)	Summary score mean (SD)	Coefficient of variation (%) (SD/mean*100)	Categorical burden (%)			
										No	Low	Medium	High
Tran, 2012 [9]	France	≥1 CC	PC & TC	502	59.3 (17.0)	53.1	0–130	31.0 (25.3)	81.6				
Tran, 2014 [20]	Multiple	≥1 CC	Online community	610	51.2 (12.4)	76.7	0–150	53.1 (31.5)	59.3				
Sav, 2016 [24]	Australia	≥1 CC	Community, PC & TC	581	57.3 (15.6)	70.0	0–150	56.2 (34.5)	61.4				
N’Goran, 2018 [45]	Switz.	≥3 CC	PC	888	72.9 (12.0)	52.0	0–130	26.8 (18.6)	69.4				
Schreiner, 2018 [46]	USA	≥2 CC	NF	74	75.4 (10.3)	74.3	0–150	37.0 (24.4)	65.9				
Chin, 2019 [22]	China	≥2 CC	PC	200	62.0 (56.0–67.0) <sup>a</sup>	55.0	0–150	16.0 (7.25–30.0) <sup>a</sup>					
Herzig, 2019 [47]	Switzerland	≥3 CC	PC	888	72.9 (12.0)	51.8	0–150	26.8 (18.6)	69.4				
Schreiner, 2019 [26]	USA	HIV & ≥2 CC	TC	103	53.2 (7.2)	35.0	0–130	22.8 (24.6)	107.9		58.0	26.0	16.0
Schreiner, 2019 [48]	USA	HIV	Community, PC & TC	103	53.2 (7.2)	35.0	0–130	22.9 (24.5)	107.0				
Tinetti, 2019 [49]	USA	≥3 CC	PC	366	75.9 (7.2)	64.2	0–150	20.9 (24.1)	115.3				
Ysraelit, 2019 [21]	Argentina	MS	TC	171	42.0 (10.0)	65.0	0–130	43.0 (29.0)	67.4				
Messi, 2020 [50]	Switzerland	≥3 CC	PC	843	73.0 (12.0)	51.7	0–130	26.8 (18.8)	70.1				
Potpara, 2020 [51]	Serbia	CVD	TC	514	64.9 (11.3)	41.1	0–170	44.9 (23.7)	52.8				
Schreiner, 2020 [52]	USA	≥2 CC	PC	74	75.4 (10.3)	74.3	0–150	39.1 (25.9)	66.2	12.2	23.0	27.0	37.8
Schreiner, 2020 [53]	USA	HIV	TC	103	53.2 (7.2)	35.0	0–130	22.9 (24.5)	107.0				
Schreiner, 2020 [54]	USA	≥1 CC	PC & TC	149	58.1 (11.9)	58.7	0–150	38.2 (31.8)	83.2		40.9	24.2	34.9
Tran, 2020 [25]	France	≥1 CC	Online community	2,413	54.8 (15.4)	54.2	0–150	41.0 (20.0–73.0) <sup>a</sup>					
Lintzeris, 2021 [55]	Australia	Opioid dependance	PC	119	44.4 (10.5)	41.2	0–150	39.4 (30.3)	76.9				
Radojicic, 2021 [56]	USA	HA	Online community	75	44.1 (–)	80.0	0–150	44.5 (–)					
								Range of mean scores:		20.9	–	56.2	
								Range of standard deviation:		18.6	–	34.5	
								Range of median scores:		16.0	–	41.0	

Abbreviations: CC, chronic condition(s); CVD, cardiovascular disease; HA, hereditary angioedema; MS, multiple sclerosis; NF, nursing facility; PC, primary care; TC, tertiary care.

<sup>a</sup> Median (quartile 1 - quartile 3).

<sup>b</sup> Of the instrument used.

influence of traditional beliefs [64]. This gap also affects BoT-MMs' appropriateness for rural dwellers, who typically experience cultural marginalization and poor access to local health systems, among other challenges [65]. Consequently, existing and newer BoT-MMs will require further developments to serve these populations meaningfully.

Existing BoT-MMs had some limitations on their development. Recall time was included in only two BoT-MMs. In the absence of recall windows, patients may provide more information, but the likelihood of error is increased too [66]. Recall times (e.g., two-to-four weeks or longer lengths) could be incorporated, ensuring they are not adding noise to the measures. Floor effects were frequently reported among BoT-MMs [9–11,20,22,26–28,35,36,39], which means that substantial proportions of individuals obtain minimum scores; thus, the true extent of their BoT cannot be accurately determined [67]. BoT-MMs developers can reduce the floor effect in future instruments by applying tools from item response theory [68]. BoT-MMs were often administered by a third person, but these responses should be self-reported [69]; assisted interviews might skew potential responses but reflect a need given the potential fragility, literacy level, or involvement of caregivers among some multimorbid patients [14]. Both self-reported and assisted approaches could be compared to ensure accurate findings [70].

Measurement properties of extant BoT-MMs were insufficiently investigated, as assessed by COSMIN standards, and this is essential as the information collected using tools without optimal measurement properties can be deemed of uncertain value. Unclear measurement properties included those infrequently assessed (e.g., responsiveness) and those assessed but either not meeting optimal standards (e.g., content validity), being inconsistent between studies (e.g., different factor structure or internal consistency), or having incomplete data to be rated (e.g., incomplete reporting on factor analysis). This has practical implications for users, for example, that most BoT-MMs should not be used for evaluating interventions (e.g., pre-post designs), that two BoT-MMs applied to the same population could reach different conclusions, or that same BoT-MMs applied to different populations could not be measuring same outcomes (or doing it but unequally well). All these issues could be addressed in future studies that include methodology and reporting as recommended by COSMIN standards [15,16].

For measurement properties, both content validity and internal consistency are essential and support provisional recommendation of an instrument according to COSMIN [15,16], but both were optimal in neither of the BoT-MMs. For existing BoT-MMs, additional content validity studies could be conducted to improve this property, which involves assessing the comprehensiveness and comprehensibility of items by patients with multimorbidity using cognitive interviews or equivalent qualitative testing [15].

Responsiveness is a poorly investigated measurement property that informs if changes seen in within-individual BoT-MMs scores after an intervention correspond to true changes in BoT of patients (smallest detectable changes), which is linked to investigating when these changes are clinically relevant (minimal important changes) [71]. Assessment of responsiveness requires longitudinal data since cross-sectional data cannot predict it [72]. Thus, optimal BoT-MMs responsiveness, found only in one BoT-MM (MTBQ), ensures accurate evaluations of an intervention effectiveness [73], making the monitoring of BoT change before and after interventions at the population level feasible while tracking patient improvement or deterioration during care.

Data from several applicative studies reported population-based summary scores, which could be used as reference norms [71]. However, the interpretability of BoT-MMs scores remains challenging since few data assigned clinical value to patients' scores. For example, a study determined that TBQ scores  $\geq 59$  points could be used to detect patients at high BoT [25], but more of these studies are needed to enhance the interpretability of BoT-MMs. Data also showed that BoT-MMs were frequently applied in populations with single conditions rather than multimorbidity, and, in such cases, the validity of findings is unclear where local BoT-MMs adaptations were not previously conducted.

#### 4.1. Recommendations for future use of BoT-MMs

This review will aid diverse stakeholders, researchers, and clinicians interested in using BoT-MMs. Accordingly, we provide two evidence-based recommendations.

First, when deciding which BoT-MM to use, consider evaluating comprehensively their characteristics (Table 2), their measurement properties (Table 3), and the interpretability of their scores (Tables A.8). Also, BoT estimates (Table 4 and Tables A.10-A.15) may help design future studies, meta-analyses, or conduct sample size estimations.

Second, consider following standards, for example, the COSMIN guidelines [16], when preparing future validation studies of BoT-MMs. Prioritize conducting evidence on characteristics of BoT-MMs that are insufficiently developed, such as handling of floor effects and incorporation of recall times, and on measurement properties identified as either insufficient or under-investigated (Table 3).

#### 4.2. Strengths and limitations

This review summarizes the evidence of BoT-MMs derived from the critical review of over 70 primary studies. Nonetheless, some limitations are discussed. First, the COSMIN evaluation of content validity is rated and graded based on information available from development and content validity studies, and it also requires subjective judgment from reviewers. Underreporting of methods or

findings among those studies may decrease their ratings, although having conducted validity procedures but not reporting them is unexpected [16]. Second, although the COSMIN methodology has extensive criteria for conducting a comprehensive assessment of instruments, it lacks specific standards for rating the interpretability of scores; nonetheless, appropriate interpretability aspects are described as suggested by COSMIN [16], for example, describing BoT estimates and available data for interpretation of numerical and categorical scores. Third, the search was limited to one database (MEDLINE). While other databases can be searched to supplement our findings, MEDLINE was deemed to provide comprehensive coverage of biomedical repositories of studies related to treatment burden and BoT-MMs.

## 5. Conclusion

This review summarizes the evidence and gaps in the development, validation, and application of eight BoT-MMs, showing that the evidence needed to inform the application of these instruments in patients with multimorbidity remains insufficiently developed, especially in terms of suitability on their development, measurement properties, interpretability of scores, and their appropriateness to use in low-resource settings.

## CRedit authorship contribution statement

**Daniel Mendoza-Quispe:** Funding acquisition, Conceptualization, Methodology, Data curation, Formal analysis, Writing – original draft, Writing – review & editing. **Silvana Perez-Leon:** Conceptualization, Methodology, Data curation, Formal analysis, Writing – original draft, Writing – review & editing. **Christopher A. Alarcon-Ruiz:** Methodology, Formal analysis, Writing – review & editing. **Andrea Gaspar:** Funding acquisition, Conceptualization, Methodology, Writing – review & editing. **María Sofía Cuba-Fuentes:** Funding acquisition, Methodology, Writing – review & editing. **Joseph R. Zunt:** Funding acquisition, Methodology, Writing – review & editing. **Victor M. Montori:** Funding acquisition, Methodology, Writing – review & editing. **Juan Carlos Bazo-Alvarez:** Funding acquisition, Conceptualization, Methodology, Visualization, Writing – review & editing. **J. Jaime Miranda:** Funding acquisition, Conceptualization, Methodology, Visualization, Writing – review & editing.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jclinepi.2023.05.013>.

## References

- [1] Skou ST, Mair FS, Fortin M, Guthrie B, Nunes BP, Miranda JJ, et al. Multimorbidity. *Nat Rev Dis Primers* 2022;8(1):1–22.
- [2] Basto-Abreu A, Barrientos-Gutierrez T, Wade AN, Melo DO, Souza ASS, Nunes BP, et al. Multimorbidity matters in low and middle-income countries. *J Multimorbidity Comorbidity* 2022;12:1–10.
- [3] Sav A, King MA, Whitty JA, Kendall E, Mcmillan SS, Kelly F, et al. Burden of treatment for chronic illness: a concept analysis and review of the literature. *Health Expect* 2015;18:312.
- [4] Demain S, Gonçalves AC, Areia C, Oliveira R, Marcos AJ, Marques A, et al. Living with, managing and minimising treatment burden in long term conditions: a systematic review of qualitative research. *PLoS One* 2015;10:e0125457.
- [5] Alsadah A, van Merode T, Alshammari R, Kleijnen J. A systematic literature review looking for the definition of treatment burden. *Heliyon* 2020;6(4):e03641.
- [6] Hurst JR, Agarwal G, van Boven JFM, Daivadanam M, Gould GS, Wan-Chun Huang E, et al. Critical review of multimorbidity outcome measures suitable for low-income and middle-income country settings: perspectives from the Global Alliance for Chronic Diseases (GACD) researchers. *BMJ Open* 2020;10(9):e037079.
- [7] Gaspar A, Mirada JJ. Burden of treatment as a measure of healthcare quality: an innovative approach to addressing global inequities in multimorbidity. *PLoS Glob Public Health* 2022;2(5):e0000484.
- [8] Boehmer KR, Pine KH, Whitman S, Organick P, Thota A, Espinoza Suarez NR, et al. Do patients with high versus low treatment and illness burden have different needs? A mixed-methods study of patients living on dialysis. *PLoS One* 2021;16:e0260914.
- [9] Tran VT, Montori VM, Eton DT, Baruch D, Falissard B, Ravaud P. Development and description of measurement properties of an instrument to assess treatment burden among patients with multiple chronic conditions. *BMC Med* 2012;10:68.
- [10] Duncan P, Murphy M, Man MS, Chaplin K, Gaunt D, Salisbury C. Development and validation of the multimorbidity treatment burden questionnaire (MTBQ). *BMJ Open* 2018;8(4):e019413.
- [11] Eton DT, Yost KJ, Lai JS, Ridgeway JL, Egginton JS, Rosedahl JK, 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* 2017;26:489–503.
- [12] Rosbach M, Andersen JS. Patient-experienced burden of treatment in patients with multimorbidity - a systematic review of qualitative data. *PLoS One* 2017;12:e0179916.
- [13] Møller A, Bissenbakker KH, Arreskov AB, Brodersen J. Specific measures of quality of life in patients with multimorbidity in primary healthcare: a systematic review on patient-reported outcome measures' adequacy of measurement. *Patient Relat Outcome Meas* 2020;11:1–10.
- [14] Sheehan OC, Leff B, Ritchie CS, Garrigues SK, Li L, Saliba D, et al. A systematic literature review of the assessment of treatment burden experienced by patients and their caregivers. *BMC Geriatr* 2019;19:262.
- [15] Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res* 2018;27:1159–70.
- [16] Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018;27:1147–57.

- [17] Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Social Res Methodol* 2007;8(1):19–32.
- [18] Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med* 2018;169:467–73.
- [19] Mendoza-Quispe DE, Perez-Leon S, Bazo-Alvarez JC, Miranda JJ. Patient Reported Outcome Measures (PROMs) of treatment burden in multimorbidity: a scoping review protocol. *OSF* 2022;1–9. <https://doi.org/10.17605/OSF.IO/FYKEJ>.
- [20] Tran VT, Harrington M, Montori VM, Barnes C, Wicks P, Ravaud P. Adaptation and validation of the Treatment Burden Questionnaire (TBQ) in English using an internet platform. *BMC Med* 2014;12(1):1–9.
- [21] Ysraelit MC, Fiol MP, Peña FV, Vanotti S, Terrasa SA, Tran VT, et al. Adaptation and validation of a Spanish version of the treatment burden questionnaire in patients with multiple sclerosis. *BMC Neurol* 2019;19(1):209.
- [22] Chin WY, Wong CKH, Ng CCW, Choi EPH, Lam CLK. Cultural adaptation and psychometric properties of the Chinese Burden of Treatment Questionnaire (C-TBQ) in primary care patients with multi-morbidity. *Fam Pract* 2019;36:657–65.
- [23] Zhang Q, Zhang K, Li M, Gu J, Li X, Li M, et al. Validity and reliability of the Mandarin version of the Treatment Burden Questionnaire among stroke patients in Mainland China. *Fam Pract* 2021;38:537–42.
- [24] Sav A, Whitty JA, McMillan SS, Kendall E, Kelly F, King MA, et al. Treatment burden and chronic illness: who is at most Risk? *Patient* 2016;9(6):559–69.
- [25] Tran VT, Montori VM, Ravaud P. Is my patient overwhelmed?: determining thresholds for acceptable burden of treatment using data from the ComPaRe e-cohort. *Mayo Clin Proc* 2020;95(3):504–12.
- [26] Schreiner N, Perazzo J, Currie J, Daly B, Weibel A. A descriptive, cross-sectional study examining treatment burden in people living with HIV. *Appl Nurs Res* 2019;46:31–6.
- [27] Dou L, Huang J, Duncan P, Guo L. Translation, cultural adaptation and validation of the Chinese Multimorbidity Treatment Burden Questionnaire(C-MTBQ): a study of older hospital patients. *Health Qual Life Outcomes* 2020;18(1):194.
- [28] Rogers EA, Yost KJ, Rosedahl JK, Linzer M, Boehm DH, Thakur A, et al. Validating the Patient Experience with Treatment and Self-Management (PETS), a patient-reported measure of treatment burden, in people with diabetes. *Patient Relat Outcome Meas* 2017;8:143–56.
- [29] Husebø AML, Morken IM, Eriksen KS, Nordfonn OK. The patient experience with treatment and self-management (PETS) questionnaire: translation and cultural adaption of the Norwegian version. *BMC Med Res Methodol* 2018;18:147.
- [30] Eton DT, Lee MK, St Sauver JL, Anderson RT. Known-groups validity and responsiveness to change of the Patient Experience with Treatment and Self-management (PETS vs. 2.0): a patient-reported measure of treatment burden. *Qual Life Res* 2020;29:3143–54.
- [31] Eton DT, Linzer M, Boehm DH, Vanderboom CE, Rogers EA, Frost MH, et al. Deriving and validating a brief measure of treatment burden to assess person-centered healthcare quality in primary care: a multi-method study. *BMC Fam Pract* 2020;21(1):221.
- [32] Lee MK, St Sauver JL, Anderson RT, Linzer M, Eton DT. Confirmatory factor analyses and differential item functioning of the patient experience with treatment and self-management (PETS vs. 2.0): a measure of treatment burden. *Patient Relat Outcome Meas* 2020;11:249–63.
- [33] Anderson RT, Eton DT, Camacho FT, Kennedy EM, Brenin CM, DeGuzman PB, et al. Impact of comorbidities and treatment burden on general well-being among women's cancer survivors. *J Patient Rep Outcomes* 2021;5(1):1–13.
- [34] Boyd CM, Wolff JL, Giovannetti E, Reider L, Weiss C, Xue Q, et al. Healthcare task difficulty among older adults with multimorbidity. *Med Care* 2014;52:S118–25.
- [35] Gibbons CJ, Kenning C, Coventry PA, Bee P, Bundy C, Fisher L, et al. Development of a multimorbidity illness perceptions scale (MULTIPLEs). *PLoS One* 2013;8(12):e81852.
- [36] Mohammed MA, Moles RJ, Hilmer SN, Kouladjian O'Donnel L, Chen TF. Development and validation of an instrument for measuring the burden of medicine on functioning and well-being: the Medication-Related Burden Quality of Life (MRB-QoL) tool. *BMJ Open* 2018;8(1):e018880.
- [37] Krska J, Morecroft CW, Rowe PH, Poole H. Measuring the impact of long-term medicines use from the patient perspective. *Int J Clin Pharm* 2014;36(4):675–8.
- [38] Krska J, Katusiime B, Corlett SA. Validation of an instrument to measure patients' experiences of medicine use: the Living with Medicines Questionnaire. *Patient Prefer Adherence* 2017;11:671–9.
- [39] Katusiime B, Corlett SA, Krska J. Development and validation of a revised instrument to measure burden of long-term medicines use: the Living with Medicines Questionnaire version 3. *Patient Relat Outcome Meas* 2018;9:155–68.
- [40] Wang Y, Krska J, Lin B, Mei Y, Katusiime B, Guo Y, et al. Cross-cultural adaptation and reliability testing of Chinese version of the living with medicines questionnaire in elderly patients with chronic diseases. *Patient Prefer Adherence* 2020;14:2477–87.
- [41] Nabergoj PP, Makovec U, Kos M, Kerec KM. Living with medicines questionnaire: first insights from Slovenia in 46th ESCP symposium on clinical pharmacy “science meets practice: towards evidence-based clinical pharmacy services.” *Int J Clin Pharm* 2018;40(1):203–317.
- [42] Zidan A, Awaisu A, Hasan S, Kheir N. The living with medicines questionnaire: translation and cultural adaptation into the Arabic context. *Value Health Reg Issues* 2016;10:36–40.
- [43] Giovannetti ER, Wolff JL, Xue QL, Weiss CO, Leff B, Boulton C, et al. Difficulty assisting with health care tasks among caregivers of multimorbid older adults. *J Gen Intern Med* 2012;27:37–44.
- [44] Wolff JL, Boyd CM. A look at person- and family-centered care among older adults: results from a National survey [corrected]. *J Gen Intern Med* 2015;30:1497–504.
- [45] N'Goran AA, Pasquier J, Deruaz-Luyet A, Burnand B, Haller DM, Neuner-Jehle S, et al. Factors associated with health literacy in multimorbid patients in primary care: a cross-sectional study in Switzerland. *BMJ Open* 2018;8(2):e018281.
- [46] Schreiner N, Schreiner S, Daly B. The association between chronic condition symptoms and treatment burden in a skilled nursing population. *J Gerontol Nurs* 2018;44(12):45–52.
- [47] Herzig L, Zeller A, Pasquier J, Streit S, Neuner-Jehle S, Excoffier S, et al. Factors associated with patients' and GPs' assessment of the burden of treatment in multimorbid patients: a cross-sectional study in primary care. *BMC Fam Pract* 2019;20(1):1–11.
- [48] Schreiner N, Perazzo J, Digenarro S, Currie J, Daly B, Weibel A. Examining the association between item specific treatment burden and adherence in people living with HIV. *West J Nurs Res* 2020;42(7):495–502.
- [49] Tinetti ME, Naik AD, Dindo L, Costello DM, Esterson J, Geda M, et al. Association of patient priorities-aligned decision-making with patient outcomes and ambulatory health care burden among older adults with multiple chronic conditions: a nonrandomized clinical trial. *JAMA Intern Med* 2019;179(12):1688–97.
- [50] Messi M, Mueller Y, Haller DM, Zeller A, Neuner-Jehle S, Streit S, et al. A cross-sectional study of Swiss ambulatory care services use by multimorbid patients in primary care in the light of the Andersen model. *BMC Fam Pract* 2020;21(1):1–10.
- [51] Potpara TS, Mihajlovic M, Zec N, Marinkovic M, Kovacevic V, Simic J, et al. Self-reported treatment burden in patients with atrial fibrillation: quantification, major determinants, and implications for integrated holistic management of the arrhythmia. *Europace* 2020;22(12):1788–97.
- [52] Schreiner N, Daly B. A pilot study exploring treatment burden in a skilled nursing population. *Rehabil Nurs* 2020;45(3):158–65.



- [53] Schreiner N, Perazzo J, Digennaro S, Burant C, Daly B, Webel A. Associations between symptom severity and treatment burden in people living with HIV. *J Adv Nurs* 2020;76:2348–58.
- [54] Schreiner N, DiGennaro S, Harwell C, Burant C, Daly B, Douglas S. Treatment burden as a predictor of self-management adherence within the primary care population. *Appl Nurs Res* 2020;54:151301.
- [55] Lintzeris N, Dunlop AJ, Haber PS, Lubman DI, Graham R, Hutchinson S, et al. Patient-reported outcomes of treatment of opioid dependence with weekly and monthly subcutaneous depot vs daily sublingual buprenorphine: a randomized clinical trial. *JAMA Netw Open* 2021;4(5):e219041.
- [56] Radojicic C, Riedl MA, Craig TJ, Best JM, Rosselli J, Hahn R, et al. Patient perspectives on the treatment burden of injectable medication for hereditary angioedema. *Allergy Asthma Proc* 2021;42(3):S4–10.
- [57] Friis K, Lasgaard M, Pedersen MH, Duncan P, Maindal HT. Health literacy, multimorbidity, and patient-perceived treatment burden in individuals with cardiovascular disease. A Danish population-based study. *Patient Educ Couns* 2019;102(10):1932–8.
- [58] Hider SL, Bucknall M, Jinks C, Cooke K, Cooke K, Desilva EE, et al. A pilot study of a nurse-led integrated care review (the INCLUDE review) for people with inflammatory rheumatological conditions in primary care: feasibility study findings. *Pilot Feasibility Stud* 2021;7(1):9.
- [59] Song MK, Paul S, Plantinga L, Henry C, Turberville-Trujillo L. Social networks of self-care and perceived treatment burden among patients on in-center hemodialysis. *Kidney Med* 2019;1(3):97–103.
- [60] Husebø AML, Dalen I, Richardson A, Bru E, Søreide JA. Factors influencing treatment burden in colorectal cancer patients undergoing curative surgery: a cross-sectional study. *Eur J Cancer Care (Engl)* 2021;30:e13437.
- [61] Ansari S, Hosseinzadeh H, Dennis S, Zwar N. Activating primary care COPD patients with multi-morbidity through tailored self-management support. *NPJ Prim Care Respir Med* 2020;30(1):12.
- [62] Zidan A, Awaisu A, El-Hajj MS, Al-Abdulla SA, Figueroa DCR, Kheir N. Medication-related burden among patients with chronic disease conditions: perspectives of patients attending non-communicable disease clinics in a primary healthcare setting in Qatar. *Pharmacy (Basel)* 2018;6(3):85.
- [63] van der Laan DM, Elders PJM, Boons CCLM, Nijpels G, Krska J, Hugtenburg JG. The impact of cardiovascular medication use on patients' daily lives: a cross-sectional study. *Int J Clin Pharm* 2018;40(2):412.
- [64] van Zyl C, Badenhorst M, Hanekom S, Heine M. Unravelling 'low-resource settings': a systematic scoping review with qualitative content analysis. *BMJ Glob Health* 2021;6(6):e005190.
- [65] Brundisini F, Giacomini M, DeJean D, Vanstone M, Winsor S, Smith A. Chronic disease patients' experiences with accessing health care in rural and remote areas: a systematic review and qualitative meta-synthesis. *Ont Health Technol Assess Ser* 2013;13(15):1.
- [66] Clarke PM, Fiebig DG, Gerdtham UG. Optimal recall length in survey design. *J Health Econ* 2008;27(5):1275–84.
- [67] Lim CR, Harris K, Dawson J, Beard DJ, Fitzpatrick R, Price AJ. Floor and ceiling effects in the OHS: an analysis of the NHS PROMs data set. *BMJ Open* 2015;5(7):e007765.
- [68] Petrillo J, Cano SJ, McLeod LD, Coon CD. Using classical test theory, item response theory, and rasch measurement theory to evaluate patient-reported outcome measures: a comparison of worked examples. *Value Health* 2015;18(1):25–34.
- [69] FDA. Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims 2009. Available at <https://www.fda.gov/media/77832/download>. Accessed January 26, 2022.
- [70] Tsakos G, Bernabé E, O'Brien K, Sheiham A, de Oliveira C. Comparison of the self-administered and interviewer-administered modes of the child-OIDP. *Health Qual Life Outcomes* 2008;6(1):1–8.
- [71] van der Willik EM, Terwee CB, Bos WJW, Hemmelder MH, Jager KJ, Zoccali C, et al. Patient-reported outcome measures (PROMs): making sense of individual PROM scores and changes in PROM scores over time. *Nephrology (Carlton)* 2021;26(5):391.
- [72] Puhan MA, Bryant D, Guyatt GH, Heels-Ansdell D, Schünemann HJ. Internal consistency reliability is a poor predictor of responsiveness. *Health Qual Life Outcomes* 2005;3(1):1–8.
- [73] Hays RD, Hadorn D. Responsiveness to change: an aspect of validity, not a separate dimension. *Qual Life Res* 1992;1:73–5.