MHJ

LUMORA JOURNALS

Multidisciplinary Healthcare Journal

Vol. 1, No. 3 (2025) · e-ISSN 3152-8775

Critical narrative review · DOI: not registered

Choosing Arabic treatment-burden measures for adults with multimorbidity in Saudi primary care: a critical narrative review

Fatema Alawad[1],*, Mohsin Ali[2], Abdulrahman Shudayyid Lafi Al Mutairi[3], Ahmed Khaled Alkhateeb[4], Ahmed Saeed Almalki[4], Alaa Yahia Alhilaly[5]

Submission date	Acceptance date	Publication date
20 March 2025	25 March 2025	31 March 2025

[1] College of Medicine, King Faisal University, Al-Ahsa, Saudi Arabia

[2] Consultant Family Medicine, King Saud bin Abdulaziz University for Health Sciences, King Abdulaziz Medical City, Jeddah, Saudi Arabia

[3] Qassim University, Saudi Arabia

[4] King Faisal University, Al-Ahsa, Saudi Arabia

[5] University of Ha'il, Saudi Arabia

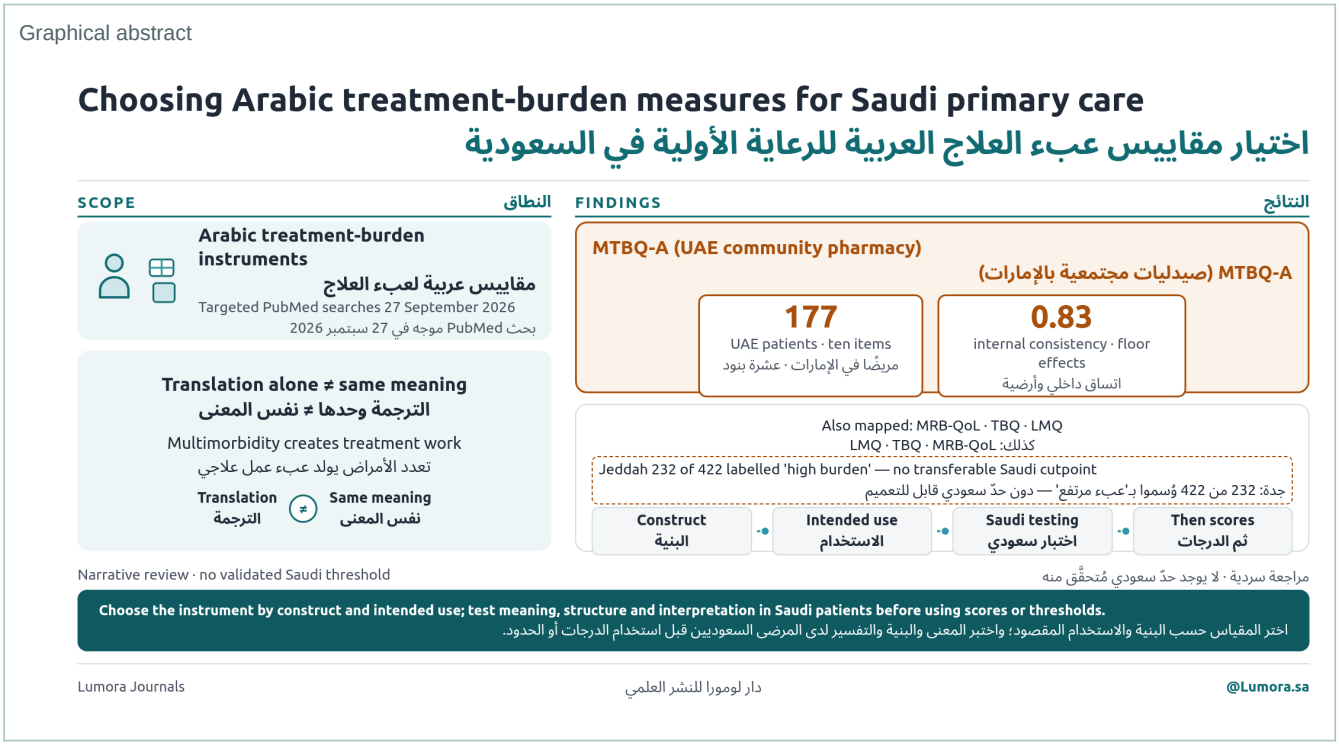
* Correspondence: Fatema Alawad, fatemalumora@gmail.com

English edition. An Arabic edition of this article is available.

Cite as: Alawad F, Ali M, Al Mutairi ASL, Alkhateeb AK, Almalki AS, Alhilaly AY. Choosing Arabic treatment-burden measures for adults with multimorbidity in Saudi primary care: a critical narrative review. Multidisciplinary Healthcare Journal. 2025;1(3):MHJ-2026-000346. © The authors. Published open access under the Creative Commons Attribution 4.0 International licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

Abstract

Graphical abstract



Background: Adults with multimorbidity perform substantial work to obtain, coordinate and use care. An Arabic translation alone does not establish the same questionnaire meaning or clinical threshold across settings. **Objective:** To distinguish Arabic treatment-burden measures and evidence needed before Saudi primary care uses scores for patient care or evaluation. **Sources and approach:** This critical narrative review used targeted PubMed searches on 27 September 2026, semantic discovery and primary-record checks. Selection was purposive, not a systematic or scoping review; no formal COSMIN rating or meta-analysis was performed. **Synthesis:** The ten-item Arabic Multimorbidity Treatment Burden Questionnaire (MTBQ-A) was tested in 177 UAE community-pharmacy patients; it had two factors, internal consistency 0.83 and substantial floor effects. Arabic Medication-Related Burden Quality of Life studies address medicines' impact rather than all multimorbidity work. An Arabic Treatment Burden Questionnaire used in Qatari renal care and an Arabic Living with Medicines Questionnaire are further antecedents, with different validation limits. A Jeddah study labelled 232 of 422 participants “high burden”, but its accessible report does not establish a transferable Saudi cutpoint or document equivalence to MTBQ-A. **Conclusion:** Instrument choice should follow construct and intended use. Saudi testing of item meaning, structure, comparability, change interpretation and clinical utility is required before scored thresholds or performance targets. Clinicians may ask patients in plain language about treatment work; scored-instrument deployment requires version, permission and local meaning checks.

Keywords: treatment burden; multimorbidity; Arabic; patient-reported outcome measure; Saudi Arabia; primary care; medicines burden

Introduction

A person with two or more long-term conditions may spend a considerable part of ordinary life arranging appointments, collecting prescriptions, monitoring symptoms, changing meals, interpreting advice from different clinicians and paying for transport or medicines. These tasks are not the same as symptoms caused by disease. Nor is their difficulty captured by a medication count or a record of missed doses. The cumulative-complexity model describes burden as an imbalance between the work that care and life

demand and the capacity available to do it; minimally disruptive medicine asks clinicians to reduce unnecessary work while pursuing outcomes that matter to the patient. [1, 2] A measure of treatment burden should therefore tell us something about *experienced work and impact*, rather than relabel adherence or disease severity.

Questionnaires can help a busy service ask about this work consistently. They can also mislead. A patient who reports little burden may have adequate support, may have stopped a demanding regimen, may interpret a translated item differently, or may be reluctant to report difficulty to a clinician. A high score may arise from a particular scheduling problem that a single total cannot identify. The 2023 review of burden-of-treatment measures in multimorbidity found eight instruments across 72 studies but described gaps in content validity, responsiveness, floor effects and score interpretation. [3] An integrative review mapped the work and impact of multimorbidity across 30 studies and noted that commonly used instruments missed the continual prioritisation that patients do when tasks compete. [4] A later systematic review of PROMs in chronic comorbidities again found sparse cross-cultural and responsiveness evidence. [5] A 2026 narrative review addressed the Treatment Burden Questionnaire in geriatric care. [6] The incremental contribution here is an **intended-use decision framework for Saudi primary care**, grounded in Arabic adaptations and the ambiguities of the Jeddah study, rather than another inventory. The added focus is a critical comparison of instrument purpose, intended use and transferability, not a claim that Arabic treatment-burden tools are absent.

There is already Arabic-language work on this topic. The Multimorbidity Treatment Burden Questionnaire has an Arabic adaptation, MTBQ-A, tested among adults with multimorbidity in the United Arab Emirates. [7] Arabic versions of a medicines-specific quality-of-life measure have progressed from content adaptation through exploratory and confirmatory testing to a hospital feasibility study. [8, 9, 10, 11] An Arabic translation of the Living with Medicines Questionnaire predates those reports. [12] A Saudi primary-care study has also reported MTBQ scores in Jeddah. [13] Consequently, a claim that Arabic measures do not exist, or that a new translation is the obvious first step, would be wrong.

The relevant Saudi question is more demanding: **When a primary-care clinic wants to learn how much work its treatment imposes on Arabic-speaking adults with multimorbidity, which measure fits the decision, which properties have been tested in comparable people, and what evidence is missing before scores guide triage or service evaluation?** This review compares constructs and settings rather than ranking instruments with a formal psychometric score. It proposes a bounded evaluation sequence, not a new scale or an implemented Saudi screening programme.

Sources and scope

Three orientation searches were run through PubMed E-utilities at 16:06 UTC on 27 September 2026, without date or language limits. The regional query was ("treatment burden"[tiab] OR "burden of treatment"[tiab]) AND (Arabic[tiab] OR Saudi[tiab]) AND (multimorbid*[tiab] OR medication*[tiab]) (seven returns). The instrument query was ("Multimorbidity Treatment Burden Questionnaire"[tiab] OR "Medication-Related Burden Quality of Life"[tiab] OR "Patient Experience with Treatment and Self-management"[tiab] OR "Treatment Burden Questionnaire"[tiab]) AND (validat*[tiab] OR psychometr*[tiab] OR translation[tiab]) (34 returns). The review query was ("treatment burden"[tiab] OR "burden of treatment"[tiab]) AND (measure*[tiab] OR instrument*[tiab]) AND (review[pt] OR "systematic review"[tiab] OR "scoping review"[tiab]) (146 returns). The exact returned identifier lists and request URLs are retained in a source log. Returns overlap; they were not all screened, deduplicated or assessed by two reviewers.

SciSpace semantic search, title/DOI lookup and references of key papers were used to locate Arabic and international antecedents that a single PubMed phrase might miss. The SciSpace results were discovery leads, not evidence until source identity and relevant claims were checked in PubMed, Crossref or an accessible primary article.

Selection was purposive: we retained primary development/adaptation studies for instruments measuring treatment work or medicines' impact; Arabic versions or applications relevant to Gulf care; the Saudi Jeddah use study; and earlier reviews that define overlap. We excluded disease symptoms, satisfaction and patient activation as burden measures, although adjacent constructs are discussed where confusion matters. No number of excluded papers is asserted because the overlapping search returns were not completely screened. The full texts of the Jeddah study, original MTBQ work, several Arabic MRB-QoL reports and selected reviews were examined through PubMed Central/Europe PMC. MTBQ-A and the 2026 Jazan qualitative report were checked at abstract level; claims do not extend beyond their accessible records. Targeted web-index checks on 27 September 2026 using site:imemr.net and site:search.emarefa.net with "treatment burden" Arabic review found no directly matching Arabic-instrument review among returned results; this is not a database-wide negative search. We did not request unavailable publisher articles, perform paired screening, formally rate risk of bias or measurement quality, calculate a pooled estimate, or claim to have found every Arabic instrument. The counts above are search returns, not a systematic-review flow. This article is a critical narrative review, not a scoping or systematic review. Table 1 makes the central source access and selection boundary visible.

Table 1. Central source families selected for this narrative decision question

Instrument or source and why selected	Population, evidence assessed and exact primary source	Access and inference boundary
MTBQ/MTBQ-A: broad multimorbidity treatment work	UK development (n=1,546) and UAE Arabic adaptation (n=177); structure, reliability, floor effects [14, 7]	UK full text; UAE abstract. The UAE two-factor and floor-effect statements are abstract-supported; neither paper validates a Saudi cutoff.
MRB-QoL Arabic: medicines' impact	UAE content adaptation, exploratory n=400, confirmatory n=343, and hospital feasibility n=227 [8, 9, 10, 11]	Available content, confirmatory and feasibility full texts; exploratory abstract. Four-factor, error and usability estimates remain source-setting specific.
Arabic LMQ and TBQ: relevant alternatives	LMQ translation/cognitive debriefing; TBQ use among 280 Qatar renal patients [12, 15, 16]	Publisher/primary records; the 2023 Qatar paper describes the Arabic TBQ translation. Renal application is not general Saudi multimorbidity validation.
Saudi Jeddah MTBQ application	Cross-sectional n=422, reported categories and adherence association [13]	Full text. Translation route, administered Arabic version and numeric category cutpoint cannot be reconstructed from the publication.
International antecedents and overlap reviews	French/English TBQ, US PETS and prior instrument/geriatric reviews [17, 18, 19, 3, 6]	Mixed full-text/abstract access; included to delimit construct and novelty, not as locally validated Arabic choices.

Table 1 footnote: This purposive map is neither an included-study denominator nor a formal COSMIN appraisal. “Access” describes what was available during this review, not all material that a journal or institution could obtain. The accompanying author research archive retains exact PMID lists, retrieval timestamp and source-level access notes; release of those records as a supplement requires accountable author approval.

What a treatment-burden score should represent

Treatment workload has multiple components. A patient must understand what to do, make the regimen compatible with work and family, reach services, manage medicines and monitoring, and repair conflicting plans. The consequences include fatigue, lost time, financial pressure, reduced social participation and strain in relationships with clinicians. [1, 19, 4] Some instruments ask mainly about *the difficulty of doing tasks*; others ask about *how medicines affect functioning and well-being*. Both can be useful, but they answer different questions. A Saudi clinic evaluating appointment coordination may learn little from a medicine-only impact score. A pharmacist reviewing a complex regimen may find that same score more actionable than a broad total.

Adherence is related yet distinct. A patient can follow a burdensome plan at great personal cost, or can have low recorded adherence because the plan is unaffordable, confusing, clinically inappropriate or simply not aligned with the patient's goals. The original Treatment Burden Questionnaire was negatively correlated with treatment satisfaction, and the English adaptation found higher burden among people reporting lower medication adherence; those associations did not make the scales interchangeable. [17, 18] A Saudi Arabic treatment-satisfaction validation in warfarin clinics measures satisfaction with anticoagulant therapy, not the full work of living with several conditions. [20] The Arabic Patient Activation Measure tested in Riyadh primary care assesses knowledge, beliefs, confidence and skills, which are relevant to capacity, but it is not a direct treatment-workload measure. [21] Choosing the familiar questionnaire because it is already translated would answer the wrong question if the intended outcome is burden.

The unit of interpretation also matters. A one-time score can start a discussion, while a change score requires evidence that the scale can detect meaningful change beyond measurement error in that setting. Comparison across education, age, nationality or clinic types requires evidence that the items function similarly across those groups. A high internal-consistency coefficient does not establish measurement invariance or responsiveness: it tells us that items covary in one sample, and may even rise when items are redundant. If a service intends to use a cutoff to trigger extra appointments, its sensitivity, false positives, missing-data behavior and consequences need testing under the intended workflow. The evidence for simply eliciting an item-level concern is less demanding than that for labelling a patient “high burden” or comparing hospitals.

Administration is part of what a score means. An interviewer can help someone who has difficulty reading, but a patient may then hesitate to disclose cost, confusion or dissatisfaction, particularly if the interviewer is a member of the treating team. A self-completed digital form reduces that interpersonal pressure for some people while excluding others who cannot access it or who need assistance. Proxy completion by a family member may capture real family labour yet should not be silently merged with the patient's own report. A Saudi validation should record mode, language variety, assistance and relationship to the clinic for every response and examine whether item distributions change across these conditions. The approach must preserve the patient's ability to describe a burden that the fixed response options did not anticipate. These are measurement-design requirements inferred from the construct and the broader Arabic PROM literature, not outcomes demonstrated by the existing Saudi MTBQ study. [22, 23]

Established instruments and the Arabic evidence

The 10-item MTBQ was developed in UK primary care using literature, patient involvement and cognitive interviews, then tested in 1,546 adults with multimorbidity whose mean age was 71. It supported a one-

factor structure and Cronbach alpha of 0.83; the authors observed expected correlations with other burden measures and quality of life and examined change over nine months. [14] This gives the original score a coherent development history, not automatic validity in Arabic-speaking or younger Saudi populations. Further UK testing in 244 adults found test–retest agreement of 0.765, yet reversing the response scale did not solve floor effects or skewness. [24] A score crowded near zero can miss change among people whose difficulties are concentrated in one neglected domain.

The MTBQ-A is an evaluated Arabic-language adaptation tested among UAE community-pharmacy patients, not a validation study of Saudi primary-care patients. Professional forward–backward translation, expert review, cognitive interviews and pilot work preceded testing in 177 adults with multimorbidity recruited through UAE community pharmacies. In the 177-participant UAE validation, the reported content-validity index was 0.94 and internal-consistency coefficient was 0.83; observed correlations supported construct validity within that source setting but not an established Saudi clinical threshold. Factor analysis supported **two** dimensions, described by the authors as self-management/social support and burden of visiting services and professionals, rather than the one-factor pattern of the original UK study. Response distributions were positively skewed with high floor effects. [7] These findings warrant use in the population studied; they do not establish that the UK global thresholds, factor scoring or sensitivity to intervention change transfer unchanged to Saudi primary care. Nor does the abstract provide a Saudi dialect or literacy analysis.

A floor effect has practical consequences beyond a psychometric label. If many people select the lowest response across most items, a later reduction in one concrete difficulty may leave the total unchanged. Conversely, a single item about service access can worsen while the total remains low. The two-factor finding also means that a clinic should not assume a single sum captures the same construct as the original UK score without testing that interpretation. Even if a locally fitted model resembles the UAE model, this would not by itself establish a clinically useful cutoff or a patient-important change. Those require anchors linked to what patients say has improved, with an explicit time frame and adequate follow-up. The later UK study's persistence of floor effects after response-scale reversal suggests that merely changing the direction of response options is unlikely to solve the underlying content or distribution problem. [24]

Broader comparisons are useful precisely because factor structure can vary across translations and settings. Danish, German and Dutch adaptations of the MTBQ each studied their own populations and measurement properties; their existence illustrates why a translated name is not a universal psychometric result. [25, 26, 27] The question for a Saudi service is whether the tasks named by the items, response options and time frame fit its own care system and patient lives. An item about travel or obtaining appointments may behave differently when care is dispersed across providers, when a family member arranges visits, or when medicines are supplied through several systems. These are hypotheses for local interviews and analysis, not findings already established by the Gulf validation.

The 31-item Medication-Related Burden Quality of Life tool, MRB-QoL, began in Australia with an emphasis on the impact of medicines on daily functioning and well-being. Its initial study in 367 respondents produced five domains, including regimen complexity, psychological and social burden, functional roles and therapeutic relationships. [28] Arabic adaptation involved expert review and patient cognitive interviews and reported a final content-validity index of 0.96. That 2024 report correctly asked for additional psychometric work before broad clinical use. [8] Subsequent UAE studies supplied some of that work: exploratory testing among 400 hospital patients found a four-factor Arabic structure, alpha 0.973,

test–retest intraclass correlation 0.994 and a reported minimum detectable change of 3.89 points under that study's conditions. [9] Confirmatory testing in 343 adults from four UAE tertiary hospitals supported first- and second-order models of the **four-domain Arabic version**; its factor structure differs from the original English five-domain pattern. [10] These are important advances, but very high reliability coefficients are not evidence that every patient will find the questions relevant or that a four-point change is a meaningful improvement in Saudi primary care.

A mixed-methods UAE study then had 10 clinical pharmacists use MRB-QoL Arabic during routine medication reviews for 227 hospital patients, with a mean System Usability Scale score of 82.2. [11] Its non-randomised, non-controlled design provides preliminary feasibility and acceptability evidence, not an estimate of better adherence, fewer admissions or reduced patient burden caused by the questionnaire. The hospital population and pharmacist-led workflow are also different from a general Saudi primary-care consultation. The appropriate next question is whether the information changes a jointly agreed plan in a way patients recognize as helpful, without adding more work.

An earlier Arabic Living with Medicines Questionnaire translation addressed medicines-related burden through forward/back translation and cognitive debriefing, with developer permission; its 2016 report called for further validation. [12] The 15-item Treatment Burden Questionnaire was also administered in Arabic to 280 patients with advanced chronic kidney disease in Qatar; a subsequent paper from that cohort describes linguistic translation, but a disease-specific application does not establish broad Saudi multimorbidity validity or a primary-care cutoff. [15, 16] The original French and English TBQ and the more extensive Patient Experience with Treatment and Self-management scale, PETS, bring broader task and impact domains but cannot be described as Saudi-validated Arabic alternatives. The French TBQ was tested in 502 patients, and its English adaptation in 610; PETS development and later work provide domain, test–retest and change evidence in US samples. [17, 18, 19, 29, 30] A 2026 short-questionnaire development report used cognitive interviews with 15 UK adults; it does not itself validate a Saudi Arabic short form. [31] Table 2 places these families against the decision they could support without suggesting that each is ready for local threshold use.

Table 2. Selected treatment-burden instruments and the limits of transfer to Saudi primary care

Instrument and construct	Evidence relevant to Arabic use	Appropriate interpretation and open issue
MTBQ-A: broad multimorbidity treatment work [14, 7, 24]	Ten-item UAE community-pharmacy validation, n=177; two factors, alpha 0.83 and high floor effects. Original UK MTBQ had one factor; reversing responses did not resolve UK floor effects.	A concise starting point for general care work. Test Saudi item meaning, factor structure, missingness and thresholds before comparing clinics or labelling a patient high burden.
MRB-QoL Arabic: medicine impact on functioning/well-being [28, 8, 9, 10, 11]	Thirty-one items; Arabic content, exploratory and confirmatory evidence in UAE samples; four Arabic factors versus five original English factors; hospital pharmacist feasibility.	Useful when medicine work and impact are the decision. UAE measurement-error and usability estimates are not Saudi primary-care change thresholds or proof of outcome benefit.
Living with Medicines Questionnaire, Arabic adaptation [12]	Translation and cognitive debriefing reported in 2016; authors called for further validation.	A relevant medicines-burden antecedent. Confirm current version, permissions and further validation before score-level use.

Instrument and construct	Evidence relevant to Arabic use	Appropriate interpretation and open issue
Arabic TBQ use in Qatar renal care [15, 16]	Fifteen-item Arabic TBQ administered to 280 advanced chronic kidney disease patients in Qatar; an associated article describes linguistic translation.	A real Arabic antecedent, but not broad Saudi multimorbidity validation or proof of primary-care cutoff utility.
TBQ/PETS: broader international antecedents [17, 18, 19, 29, 30]	French/English TBQ and US PETS studies report development, reliability and some responsiveness; these are construct comparators, not ready Saudi Arabic alternatives.	Consider extra domains only if essential to the decision and locally adapted/validated; do not import English cutoffs.
Short Treatment Burden Questionnaire [31]	UK cognitive-interview development with 15 adults; not a completed Saudi Arabic validation.	Possible future low-workload design, not a ready local screening threshold.

Table 2 footnote: The rows distinguish instrument purpose and source populations; they are not a formal COSMIN ranking or an exhaustive register of Arabic tools. “No comparable Saudi validation found” describes targeted searches, not proof of absence. Factor counts, coefficients and sample sizes refer to cited studies and should not be pooled. Use and translation rights must be checked with each instrument owner before local deployment.

What the Saudi observations do and do not establish

A Jeddah cross-sectional study recruited 422 people with at least two clinically diagnosed long-term conditions from 12 primary-care centres in late 2023. It administered the MTBQ alongside a medication-adherence scale and labelled 232/422 (55%) “high burden”, with lower adherence scores in that category. [13] This is the source's label, not a national prevalence estimate, intervention effect or validated Saudi triage rule. The accessible full paper cites the original UK MTBQ but does not specify a numeric high/medium/low scoring rule, the exact Arabic wording administered, use or equivalence of the separate UAE MTBQ-A, or a local cognitive-testing route. Its stated eligibility is age 18 years or older while Table 1 reports an age range of 15–88; its occupation significance statement also conflicts with the printed table *P* value. It excluded several groups, including patients in palliative care, active cancer and selected mental health diagnoses, and did not perform multivariable regression because assumptions failed in small categories. These inconsistencies do not negate the participants' reported difficulties, but require source-specific caution and preclude transferring its category or prevalence estimate to other clinics.

The numerical category is particularly sensitive to the scoring rule. The original MTBQ investigators grouped nonzero scores into sample tertiles for interpretation; this is a research convention, not a universal biological boundary. [14] A later clinic that imports “high” as a fixed clinical label without checking the Jeddah scoring procedure, version, translation and case mix could classify people differently for reasons unrelated to their experience. The Jeddah study reports a cross-sectional association between burden and adherence, but does not determine whether demanding treatment reduces adherence, whether patients struggling with adherence perceive greater burden, or whether social circumstances influence both. The fact that active cancer and some mental health conditions were excluded further narrows transfer to a general multimorbidity clinic. A Saudi service should therefore retain raw item and domain patterns alongside any exploratory category and state exactly who was eligible to complete the form.

The 2026 Jazan qualitative report interviewed 19 adults taking at least five long-term medicines and described information work, self-devised medication routines, psychological strain and the relevance of Ramadan. [32] Only the indexed abstract was available for this review. It helps frame possible interview

prompts and contextual hypotheses; it cannot validate a questionnaire, quantify regional burden or prove that a particular item has different meaning during Ramadan. Local patient participation must determine whether and how such topics belong in the measure or clinical conversation. A medicine count by itself remains a poor substitute for that account of effort and capacity.

These observations also expose a practical tension. A broad scale may make service access visible while overlooking a patient whose main difficulty is the adverse effect or cost of one medicine. A medicine-focused scale can uncover that problem yet leave appointment travel, conflicting advice and self-monitoring outside its score. A clinic may need one brief general opening question and a more detailed domain probe only when relevant. This is a design hypothesis, not a validated combination of instruments. Repeated questionnaire completion itself is treatment work; the service should show that it acts on responses before asking patients to supply more data. The 2022 and 2023 reviews of Arabic PROMs reported that cross-cultural validity, responsiveness and measurement error have often been underexamined across fields, reinforcing the need to test the intended *use*, not merely translate more items. [22, 23]

From score to conversation, evaluation and governance

For an individual consultation, a clinician may ask open, plain-language questions about which care task or effect is most difficult, invite correction and agree whether a safe change in the care plan is possible. This does not require administering, copying or scoring a proprietary questionnaire. Using a published item's actual wording, even informally, requires a version and permission check; a reproducible score also requires local meaning and workflow testing. If a locally authorized instrument is later used, a low total should not end the conversation when a single item signals difficulty, and a high total should not automatically trigger extra appointments or medicine changes. The patient may want fewer visits, a clearer plan, home delivery, a different dosing schedule or help coordinating specialists; each action has its own clinical safety context. A questionnaire is not itself a diagnosis, an adherence judgment or a treatment order. [1, 2, 19]

For service evaluation, the threshold is higher. If a clinic uses MTBQ-A to compare sites or dates, it needs a defined eligible population, consistent administration mode and language, recorded nonresponse, and evidence that scores mean comparable things across groups. It must also prespecify case-mix variables and describe who can and cannot complete the form. Differences may reflect disease severity and mix, selection into care or survey response, administration mode and the services available, rather than clinic quality. If a pharmacist uses MRB-QoL Arabic before and after a medicine review, measurement error and responsiveness must be tested in that Saudi setting; its UAE minimum detectable change of 3.89 is a study-specific error estimate, **not** a patient-important improvement threshold. [9] The output should sit beside balancing outcomes: medicine safety, visits, cost, time spent completing questionnaires and patient-reported benefit. Collecting forms while leaving care unchanged may increase the burden being measured.

Table 3 states a proposed evaluation sequence. The COSMIN PROM framework distinguishes content and structural validity, reliability, measurement error, responsiveness and interpretability; this review uses those **domains** to explain why intended uses demand different evidence but does not conduct a COSMIN review or rating. [33] The local interview, workflow, case-mix and safety steps in Table 3 are the review authors' proposals, not published COSMIN requirements, nor a checklist endorsed by MHJ or a Saudi clinical society. The sequence should be designed with Arabic-speaking patients of varying education, age, disease combinations and care arrangements, including those who need help reading. Researchers should specify whether the goal is a conversation, group comparison, change detection or triage *before* choosing

an instrument. That choice determines the indispensable measurement evidence and whether an extensive scale is justified in a short consultation.

Table 3. Proposed evidence needed for four intended uses in Saudi primary care

Intended use	Evidence to obtain for that use	Consequence to monitor
Open a patient conversation [7, 8]	Plain-language inquiry needs no score. If a questionnaire is administered: verify current version/permission, test local item meaning, acceptability, completion time, assistance and missingness.	Whether a concern leads to an agreed, safe change without adding work.
Compare patients, groups or clinics [14, 10, 3, 33]	Local structure, reliability and item-functioning evidence; consistent administration; explicit eligibility and nonresponse denominators; prespecified case-mix and mode adjustment or stratification.	Unfair comparisons arising from age, illness mix, literacy, service access, site or response selection.
Track change after a service intervention [29, 24, 9]	Test–retest, measurement error, responsiveness, patient-anchored interpretation and timing appropriate to the intervention; do not transplant UAE/UK thresholds.	Clinical harm, extra contacts, total time spent in care and whether patients value the change.
Use a cutoff to triage support [13, 5]	Prospectively validate the cutoff against a defined action and outcomes, including false positives/negatives and capacity to provide the promised response.	A high-score label that stigmatizes patients or increases visits without reducing burden.

Table 3 footnote: Content/structural validity, reliability, error and responsiveness are recognized measurement domains in COSMIN guidance [33]; the specified Saudi workflow, case-mix, safety and clinical-impact steps are the review authors' proposals. This table supplies no validated Saudi thresholds, individualized rules or formal COSMIN rating. Clinical, information-governance and patient representatives must approve any implemented workflow and confirm instrument-use terms.

The study design should separate measurement testing from clinical-effect testing. First, qualitative interviews can reveal concepts the selected instrument misses and items whose interpretation depends on clinic or family context. A small pilot can then measure completion burden and missingness. A larger prospective validation can test structure, reliability and cross-group comparability; a repeated-measures cohort with a prespecified anchor can examine change. Only after the score has a credible interpretation should a controlled or carefully designed pragmatic evaluation ask whether feeding results back to clinicians improves a patient-important outcome. Training, time to review scores and a route to act on problems are part of the intervention, not background assumptions. The UAE pharmacist feasibility study demonstrates one way to examine usability, but it neither substitutes for Saudi transport testing nor proves benefit. [11]

For that final evaluation, the comparison should be ordinary care with the same opportunity for patients to raise concerns, not a group denied necessary clinical attention. Teams could predefine actions after an item response—for example, a medicine-review referral for a regimen problem or a scheduling review for repeated visits—while leaving individualized decisions to the treating clinicians. Outcomes should include the patient's own account of whether work became more manageable, completed changes to the care plan, safety events, unplanned care and the time patients and staff spent on the new process. A process measure such as “forms completed” is insufficient: it can improve even as unresolved burden grows. Missing forms matter too, because the people with least capacity may be least likely to complete them.

Interpretation should prespecify how those missing responses are described and avoid treating them as zero burden. The goal of this sequence is to determine whether measurement changes care in a useful and safe direction, rather than to reward a high response rate alone.

Rights and accountability are part of the evaluation. The Living with Medicines Questionnaire translation reported obtaining developer permission; other instruments may have different access and reproduction terms. [12] Any clinic intending to copy items into an electronic record or translate them further must verify the current instrument version and use terms. Data collection also requires a responsible clinical team, a plan for responding to urgent safety concerns, and appropriate institutional information-governance decisions. This review uses published studies and no new patient data; it cannot supply approvals or consent for a future validation study.

Interpretation and limitations

This review makes a narrower claim than a psychometric synthesis. It does not formally rate the quality of each study, establish measurement invariance, recommend one instrument for every setting, or determine a Saudi score threshold. The source search was targeted and may have missed non-indexed Arabic work. Some pivotal records, including MTBQ-A and the Jazan interview study, were assessed from abstracts; the UAE MRB-QoL and Jeddah full texts were examined more closely. This uneven access limits fine-grained judgments about sampling, missing data and item performance. The broad systematic and scoping reviews cited here have their own populations and time windows; their counts cannot be treated as a fresh search completed for this article. [3, 22, 23, 5]

The contrast between MTBQ-A and MRB-QoL Arabic is primarily a contrast of *construct and setting*, not a head-to-head performance comparison. No identified primary study administered both in the same Saudi primary-care sample. The Jeddah study's “high burden” category is its own reported category, not evidence that the UK or UAE cutoff is ready for universal Saudi triage. The strongest near-term conclusion is that there are Arabic tools worth evaluating, while the strongest open question is how they behave and affect care when used in the intended Saudi service. No identified Saudi primary-care study directly compared these instruments or established that applying their scores improves clinical decision-making.

Conclusion

Arabic treatment-burden measurement should begin with the decision it is meant to inform. MTBQ-A offers a brief view of multimorbidity care work; MRB-QoL Arabic explores medicines' impact in greater detail; Arabic LMQ and TBQ applications provide other antecedents with different limits. A Saudi clinician can ask open questions in plain language about what care costs a patient in time, effort and opportunity. A clinic should administer and score a named questionnaire only after checking its current version, permission and local meaning; it should not use an imported cutoff as a diagnostic label or quality target until comparability, change interpretation and clinical consequences have been tested. A patient-centred validation and implementation study is a more defensible next step than another uncontextualized prevalence percentage.

Declarations

Ethics: Ethical approval and informed consent were not required because this review uses only published literature and no new participant data were collected.

Funding: This research received no external funding.

Competing interests: The authors declare no competing interests.

Data availability: The targeted PubMed queries are reported in the text; search logs and source-selection notes are available from the corresponding author on request. Citation of measures does not grant permission to reproduce their items or use scoring software; prospective users should confirm current terms with instrument owners.

References

1. Shippee ND, Shah ND, May CR, Mair FS, Montori VM. Cumulative complexity: a functional, patient-centered model of patient complexity can improve research and practice. J Clin Epidemiol. 2012;65(10):1041-51. doi:10.1016/j.jclinepi.2012.05.005. https://pubmed.ncbi.nlm.nih.gov/22910536/

2. Abu Dabrh AM, Gallacher K, Boehmer KR, Hargraves IG, Mair FS. Minimally disruptive medicine: the evidence and conceptual progress supporting a new era of healthcare. J R Coll Physicians Edinb. 2015;45(2):114-7. doi:10.4997/JRCPE.2015.205. https://pubmed.ncbi.nlm.nih.gov/26181525/

3. Mendoza-Quispe D, Perez-Leon S, Alarcon-Ruiz CA, Gaspar A, Cuba-Fuentes MS, Zunt JR, et al. Scoping review of measures of treatment burden in patients with multimorbidity: advancements and current gaps. J Clin Epidemiol. 2023;159:92-105. doi:10.1016/j.jclinepi.2023.05.013. https://pubmed.ncbi.nlm.nih.gov/37217106/

4. Lee JE, Lee J, Shin R, Oh O, Lee KS. Treatment burden in multimorbidity: an integrative review. BMC Prim Care. 2024;25(1):352. doi:10.1186/s12875-024-02586-z. https://pubmed.ncbi.nlm.nih.gov/39342121/

5. Wang Y, Li CY, Zhou JT, Liu WY, Zhu ZY, Zheng FS. Patient-reported outcome measures for patients with chronic comorbidities: a systematic review of psychometric properties. Qual Life Res. 2026;35(1):23. doi:10.1007/s11136-025-04136-1. https://pubmed.ncbi.nlm.nih.gov/41484477/

6. Li-Wang J, Hernandez S, Ramasamy S, Qu L, Banks J, Samper-Ternent R, et al. Measuring treatment burden with the treatment burden questionnaire (TBQ): a narrative literature review and implications for geriatric research and clinical practice. Support Care Cancer. 2026;34(6):552. doi:10.1007/s00520-026-10798-5. https://pubmed.ncbi.nlm.nih.gov/42149265/

7. Samorinha C, Saidawi W, Duncan P, Alzoubi KH, Alzubaidi H. Translation, cross-cultural adaptation and validation of the Arabic multimorbidity treatment burden questionnaire (MTBQ-A): A study of adults with multimorbidity. Res Social Adm Pharm. 2024;20(4):411-418. doi:10.1016/j.sapharm.2024.01.001. https://pubmed.ncbi.nlm.nih.gov/38267315/

8. Al-Ebrahim SQ, Harrison J, Chen TF, Alzubaidi H, Mohammed MA. The Arabic medication-related burden quality of life (MRB-QoL) tool: Cross-cultural adaptation and content validation. Explor Res Clin Soc Pharm. 2024;16:100523. doi:10.1016/j.rcsop.2024.100523. https://pubmed.ncbi.nlm.nih.gov/39498226/

9. Al-Ebrahim SQ, Hafidh K, Harrison J, Chen TF, O'Donnell LK, AlHusseini A, et al. The medication-related burden quality of life (MRB-QoL) Arabic tool: Exploratory factor analysis and psychometric evaluation. Res Social Adm Pharm. 2025;21(8):608-619. doi:10.1016/j.sapharm.2025.04.001. https://pubmed.ncbi.nlm.nih.gov/40300964/

10. Al-Ebrahim SQ, Hafidh K, Jallo M, Mauwfak MM, Nassef M, Alzubaidi H, et al. The Medication-Related Burden Quality of Life (MRB-QoL) tool: A confirmatory factor analysis of the Arabic version. Explor Res Clin Soc Pharm. 2025;17:100568. doi:10.1016/j.rcsop.2025.100568. https://pubmed.ncbi.nlm.nih.gov/39926417/

11. Al-Ebrahim SQ, El Ouweini A, Boura F, Abu Tayyem HM, Diab R, Adas O, et al. Clinical utility of the Arabic medication-related burden quality of life (MRB-QoL) tool in hospital-based medicines optimisation services: A mixed methods feasibility study. Explor Res Clin Soc Pharm. 2025;19:100620. doi:10.1016/j.rcsop.2025.100620. https://pubmed.ncbi.nlm.nih.gov/40585785/

12. 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. doi:10.1016/j.vhri.2016.07.001. https://pubmed.ncbi.nlm.nih.gov/27881275/

13. Khojah YY, Bashawri MA, Khojah NY, Hassanien NSM. Prevalence of and Factors Associated with Treatment Burden and Medication Adherence Among Primary Care Patients with Multimorbidity in Jeddah, Saudi Arabia: A Cross-Sectional Study. Cureus. 2024;16(8):e67514. doi:10.7759/cureus.67514. https://pubmed.ncbi.nlm.nih.gov/39310460/

14. 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. doi:10.1136/bmjopen-2017-019413. https://pubmed.ncbi.nlm.nih.gov/29654011/

15. Al-Mansouri A, Al-Ali FS, Hamad AI, Mohamed Ibrahim MI, Kheir N, Ibrahim RA, et al. Assessment of treatment burden and its impact on quality of life in dialysis-dependent and pre-dialysis chronic kidney disease patients. Res Social Adm Pharm. 2021;17(11):1937-1944. doi:10.1016/j.sapharm.2021.02.010. https://pubmed.ncbi.nlm.nih.gov/33612446/

16. Al-Mansouri A, Hamad AI, Al-Ali FS, Ibrahim MIM, Kheir N, Al-Ziftawi NH, et al. Pill-burden and its association with treatment burden among patients with advanced stages of chronic kidney disease. Saudi Pharm J. 2023;31(5):678-686. doi:10.1016/j.jsps.2023.03.008. https://pubmed.ncbi.nlm.nih.gov/37181136/

17. 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. doi:10.1186/1741-7015-10-68. https://pubmed.ncbi.nlm.nih.gov/22762722/

18. 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;12:109. doi:10.1186/1741-7015-12-109. https://pubmed.ncbi.nlm.nih.gov/24989988/

19. 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(2):489-503. doi:10.1007/s11136-016-1397-0. https://pubmed.ncbi.nlm.nih.gov/27566732/

20. Shilbayeh SAR, Alyahya SA, Alshammari NH, Almutairi WA, Shaheen E. Treatment Satisfaction Questionnaire for Medication: Validation of the Translated Arabic Version among Patients Undergoing Warfarin Therapy in Saudi Arabia. Value Health Reg Issues. 2018;16:14-21. doi:10.1016/j.vhri.2018.01.007. https://pubmed.ncbi.nlm.nih.gov/29626737/

21. Kerari A, Almalki M, Bahari G, Alharbi MF. Validation of the Arabic Version of the Patient Activation Measure (PAM-13) for Application within the Primary Healthcare Context in Saudi Arabia. Healthcare (Basel). 2023;11(23):3090. doi:10.3390/healthcare11233090. https://pubmed.ncbi.nlm.nih.gov/38063658/

22. Alaqeel S, Alfakhri A, Alkherb Z, Almeshal N. Patient-reported outcome measures in Arabic-speaking populations: a systematic review. Qual Life Res. 2022;31(5):1309-1320. doi:10.1007/s11136-021-03010-0. https://pubmed.ncbi.nlm.nih.gov/34709559/

23. Al-Ebrahim SQ, Harrison J, Chen TF, Mohammed MA. Cross-cultural adaptation and psychometric properties of patient-reported outcome measures in Arabic speaking countries: A scoping review. Res Social Adm Pharm. 2023;19(7):989-1006. doi:10.1016/j.sapharm.2023.03.007. https://pubmed.ncbi.nlm.nih.gov/36941158/

24. Duncan P, Scott LJ, Dawson S, Munas M, Pyne Y, Chaplin K, et al. Further development and validation of the Multimorbidity Treatment Burden Questionnaire (MTBQ). BMJ Open. 2024;14(4):e080096. doi:10.1136/bmjopen-2023-080096. https://pubmed.ncbi.nlm.nih.gov/38604632/

25. Pedersen MH, Duncan P, Lasgaard M, Friis K, Salisbury C, Breinholt Larsen F. Danish validation of the Multimorbidity Treatment Burden Questionnaire (MTBQ) and findings from a population health survey: a mixed-methods study. BMJ Open. 2022;12(1):e055276. doi:10.1136/bmjopen-2021-055276. https://pubmed.ncbi.nlm.nih.gov/34980626/

26. Schulze J, Breckner A, Duncan P, Scherer M, Pohontsch NJ, Lühmann D. Adaptation and validation of a German version of the Multimorbidity Treatment Burden Questionnaire. Health Qual Life Outcomes. 2022;20(1):90. doi:10.1186/s12955-022-01993-z. https://pubmed.ncbi.nlm.nih.gov/35658972/

27. Engels LWS, van Merode T, Heijmans M, Menting J, Duncan P, Rademakers J. Measurement of treatment burden in patients with multimorbidity in the Netherlands: translation and validation of the Multimorbidity Treatment Burden Questionnaire (NL-MTBQ). Fam Pract. 2024;41(6):901-908. doi:10.1093/fampra/cmad100. https://pubmed.ncbi.nlm.nih.gov/37878729/

28. 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. doi:10.1136/bmjopen-2017-018880. https://pubmed.ncbi.nlm.nih.gov/29330175/

29. 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(11):3143-3154. doi:10.1007/s11136-020-02546-x. https://pubmed.ncbi.nlm.nih.gov/32524346/

30. 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;11:249-263. doi:10.2147/PROM.S282728. https://pubmed.ncbi.nlm.nih.gov/33447118/

31. Goulding R, Duncan P, Kovalenko AG, Gamlin C, Chilcott S, Fraser SDS, et al. Development of the Short Treatment Burden Questionnaire (STBQ) including a Global Treatment Burden Question (GTBQ): Cognitive interviews with adults living with multimorbidity. J Multimorb Comorb. 2026;16:26335565261417393. doi:10.1177/26335565261417393. https://pubmed.ncbi.nlm.nih.gov/41767163/

32. Adawi MM, Ghazwani MM, Alharbi MF, Alabdullatif MH, Mutawwam RA, Mawkili SH, et al. Exploring the Lived Experience of Polypharmacy Among Patients in Jazan, Saudi Arabia: An Interpretive Description Study. Qual Health Res. 2026 [online ahead of print]. doi:10.1177/10497323261447190. https://pubmed.ncbi.nlm.nih.gov/42224102/

33. Mokkink LB, Elsman EBM, Terwee CB. COSMIN guideline for systematic reviews of patient-reported outcome measures version 2.0. Qual Life Res. 2024;33(11):2929-2939. doi:10.1007/s11136-024-03761-6. https://pubmed.ncbi.nlm.nih.gov/39198348/

Multidisciplinary Healthcare Journal · Vol. 1, No. 3 (2025) · CC BY 4.0 · https://creativecommons.org/licenses/by/4.0/