

Development, woman-centricity and psychometric properties of maternity patient-reported experience measures: a systematic review



Claudia Bull, PhD BNutr(Hons); Alayna Carrandi, MPH, BSci;
Valerie Slavin, PhD, M Adv Prac Health Care Research, BSc Mid, Dip HE Mid, RM;
Helena Teede, AM PhD MBBS, FRACP, FAAHMS, FRANZCOG; Emily J. Callander, PhD, BA

Introduction

Women's experiences of maternity care are varied and complex. A 2020 study defined the concept of women's experiences of maternity care as "women's interpretation of their care encounters within maternity services. It is subjective in nature and evolves throughout the course of pregnancy, childbirth, and the postpartum period."¹ (page 423) The authors also highlight how a woman's needs, preferences, and expectations can influence how they perceive their maternity care.¹

Within the broader context of health service and system performance measurement, patient-reported experience

Cite this article as: Bull C, Carrandi A, Slavin V, et al. Development, woman-centricity and psychometric properties of maternity patient-reported experience measures: a systematic review. *Am J Obstet Gynecol MFM* 2023;5:101102.

From the Monash Centre for Health Research and Implementation, School of Public Health and Preventive Medicine, Monash University, Clayton, Australia (Dr Bull, Ms Carrandi, Drs Teede and Callander); Women-Newborn-Children's Services, Gold Coast Health, Southport, Australia (Dr Slavin).

Received March 29, 2023; revised July 6, 2023; accepted July 14, 2023.

The authors report no conflict of interest.

No funding supported this research.

Corresponding author: Claudia Bull, PhD BNutr (Hons). claudiabull06@gmail.com

2589-9333/\$36.00

© 2023 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>) <http://dx.doi.org/10.1016/j.ajogmf.2023.101102>

OBJECTIVE: Valid and reliable maternity patient-reported experience measures are critical to understanding women's experiences of care. They can support clinical practice, health service and system performance measurement, and research. The aim of this review is to identify and critically appraise the risk of bias, woman-centricity (content validity), and psychometric properties of maternity patient-reported experience measures published in the scientific literature.

DATA SOURCES: MEDLINE, CINAHL Plus, PsycINFO, and Embase were systematically searched for relevant records between January 1, 2010 and July 10, 2021.

STUDY ELIGIBILITY CRITERIA: We searched for articles describing the instrument development of maternity patient-reported experience measures and measurement properties associated with instrument validity and reliability testing. Articles that described patient-reported experience measures developed outside of the maternity context and articles that did not contribute to the instruments' development, content validation, and/or psychometric evaluation were excluded.

METHODS: Included articles underwent risk of bias, content validity, and psychometric properties assessments in line with the COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) guidance. Patient-reported experience measure results were summarized according to language subgroups. An overall recommendation for use was determined for each patient-reported experience measure language subgroup.

RESULTS: A total of 54 studies reported on the development and psychometric evaluation of 25 maternity patient-reported experience measures, grouped into 45 language subgroups. The quality of evidence underpinning the instruments' development was generally poor. Only 2 (4.4%) patient-reported experience measures reported sufficient content validity, and only 1 (2.2%) received a level "A" recommendation, required for real-world use.

CONCLUSION: Maternity patient-reported experience measures demonstrated poor-quality evidence for their measurement properties and insufficient detail about content validity. Future maternity patient-reported experience measure development needs to prioritize women's involvement in deciding what is relevant, comprehensive, and comprehensible to measure. Improving the content validity of maternity patient-reported experience measures will improve overall validity and reliability and facilitate real-world practice improvements. Standardized patient-reported experience measure implementation also needs to be prioritized to support advancements in clinical practice for women.

Key words: maternity, patient-reported experience measures, reliability, validity, woman-centered care

measures (PREMs) provide information on "what" happened during a care encounter and "how" it happened from the service user's perspective.² This differs from patient-reported outcome measures (PROMs), which are widely

recognized as providing information about "any aspect of a patient's health status that comes directly from the patient (ie, without the interpretation of the patient's responses by a physician or anyone else)."³ (page 1) PREMs function

AJOG MFM at a Glance

Why was this study conducted?

For patient-reported experience measures (PREMs) to meaningfully inform service improvements, they must be rigorously developed and designed for the context in which they are administered. This study identified and critically appraised maternity PREMs published in the scientific literature.

Key findings

A total of 25 maternity PREMs, comprising 45 language subgroups, were identified in the scientific literature. Only 1 PREM demonstrated sufficient evidence to receive a level “A” recommendation, required for real-world use. Other PREMs were lacking robust and consistent evidence for their content validity, psychometric properties, and quality of evidence.

What does this add to what is known?

Currently available PREMs lack content validity; in other words, they are not woman-centered. Women must be involved in the development of maternity PREMs, particularly in decisions regarding what is relevant, comprehensive, and comprehensible to measure.

as an indicator of healthcare quality alongside other measures such as patient safety and clinical effectiveness.^{4,5} PREMs also provide information about how well services align with what service users want, thereby supporting healthcare systems in providing person-centered care.^{6,7} For PREMs to meaningfully inform service improvements, they must be rigorously developed and designed for the context in which they are administered. Thus, PREMs designed for use in maternity care should best align with what women want from maternity care.

Woman-centered care (WCC) is a concept traditionally entrenched within a midwifery philosophy of care.⁸ However, it is widely recognized that WCC is the right of every woman, irrespective of where or by whom she receives care.^{9,10} One of the core tenets of WCC is that care should focus on a woman's needs, aspirations, and expectations—the concepts that PREMs intend to capture—rather than the needs of an organization or care providers.¹¹ However, a 2021 review found that no large-scale national PREM programs currently being used for maternity care actually involved women in their design of PREM surveys or in deciding what is important and relevant to measuring maternity care experiences.¹² This suggests that some maternity PREMs may not be capturing what matters most to

women, and resultantly are not woman-centric. This is a particular failure given the potential use of PREMs for service evaluation and improvement. Specifically, maternity PREMs can support care providers to be more woman-centered in their approach to care by: identifying women's preferences for involvement in their maternity care and decision-making; responding to women's desire for information such as options for pain relief; and establishing dialogue with women about challenges that they are experiencing in relation to the continuity of their maternity care.¹³

Objective

A 2021 systematic review conducted by Beecher et al¹² evaluated the methodological and psychometric quality of maternity PREMs that were designed for national surveying efforts. This review did not include PREMs published in the scientific literature, which support international research, quality improvement initiatives, and service evaluation. To address this gap, the aim of this review was to identify and critically appraise the risk of bias, woman-centricity (content validity), and psychometric properties of maternity PREMs published in the scientific literature. This review, in addition to the findings of Beecher et al,¹² will support a holistic understanding of the quality

of maternity PREMs and offer recommendations to reinforce woman-centered service and system performance measurement, activities that use PREMs.

Methods**Design and registration**

This review was conducted in line with the COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) guidance for systematic reviews of outcome measurement instruments.^{14–16} This guidance was developed for reviews related to PROMs, and there is no COSMIN equivalent for evaluating the quality of PREMs. However, because COSMIN guidance has been used to support the evaluation of PREMs elsewhere,^{2,12,17} it was deemed the most suitable and comprehensive guidance for the current review. A protocol for this systematic review was published a priori⁹ and registered on PROSPERO (International Prospective Register of Systematic Reviews; registration number: CRD42021288854).

Eligibility criteria, information sources, and search strategy

A systematic literature search was conducted to identify relevant studies published between January 1, 2020 and July 10, 2021 on the following electronic databases: MEDLINE (via Ovid), CINAHL Plus (via EBSCOhost), PsycINFO (via Ovid), and Embase (via Elsevier). We searched for articles that related to the following search terms and associated derivatives: (1) maternity care; (2) PREMs and instruments related to women's experiences of maternity care; (3) instrument development; and (4) measurement properties associated with instrument validity and reliability testing. Specifically, we used the search terms for measurement properties recommended by COSMIN.¹⁸ Appendix 1 includes the exact search strategies for each database.

Data screening

Studies were imported into the Covidence systematic review management software (Veritas Health Innovation, Melbourne, Australia),¹⁹ where duplicates were automatically removed. Two

TABLE 1
Eligibility criteria for studies reporting on maternity patient-reported experience measures

Inclusion criteria	Exclusion criteria
<ul style="list-style-type: none"> • Published between January 1, 2010 and July 10, 2021, representing contemporary instruments;^a • Published in English; • Available in full-text; and • Studies describing the development, content validation, and/or psychometric evaluation of PREMs relevant to all women receiving maternity care. 	<ul style="list-style-type: none"> • Literature reviews, meta-reviews, protocols, theses, or quality improvement activities; • PREM was used as an outcome measure (eg, in a cross-sectional study), but study did not contribute to the instruments' development, content validation, and/or psychometric evaluation; • Proxy-reported PREMs (ie, not self-reported by women); • Satisfaction or expectation measures;^b • PREMs originally developed in a context other than maternity; and • PREMs specific to only certain maternal subpopulations (eg, termination of pregnancy care).

PREM, patient-reported experience measure.

^a We included articles published before the January 1, 2010 cutoff date if they provided evidence to support the development and/or psychometric evaluation of maternity PREMs identified after January 1, 2010 to ensure that we were reporting a holistic representation of the instruments' quality; ^b The protocol publication⁹ includes a detailed explanation of this exclusion criterion.

Bull. Systematic review of maternity patient-reported experience measures. *Am J Obstet Gynecol MFM* 2023.

authors CB and AC independently screened titles and abstracts and then full-text articles using the eligibility criteria presented in Table 1. Disagreements were resolved through discussion until consensus was reached; adjudication by a third author was not necessary. The reference lists of all included studies were subsequently hand-searched CB and AC for other relevant studies. The inclusion of additional studies was based on review and consensus by both CB and AC.

Figure 1 depicts the overarching processes taken for this systematic review.

Data extraction

The following data were extracted from each included study: study authors, publication year, country where research was conducted, PREM name, PREM language, subscales/constructs captured, sample, setting, mode of administration, time point of PREM administration during maternity care continuum, recall period, number of PREM items, response options, score range, and average completion time. A summary of extracted information is presented in Appendix 2.

The evaluations described in the following sections were undertaken by CB and AC. Each author evaluated 50% of the included studies and cross-checked the other's assessments,

discussing discrepancies and reaching consensus on final scores/ratings.

Evaluating the methodological quality of the included studies

Methodological rigor of each study was evaluated using the COSMIN Risk of Bias checklist.¹⁴ The checklist comprises 10 boxes relating to instrument development, content validity, structural validity, internal consistency, cross-cultural validity/measurement invariance, reliability, measurement error, criterion validity, hypotheses testing for construct validity, and responsiveness. Only relevant boxes were completed and assigned an overall score of "very good" (highest score), "adequate", "doubtful", or "inadequate" (lowest score) on the basis of criteria specific to each box.¹⁴ Per COSMIN guidance, a "worst score counts" principle was used (ie, if a box comprising 8 criteria scores "very good" for 7 criteria, but "inadequate" for 1 criterion, the overall box score is "inadequate").

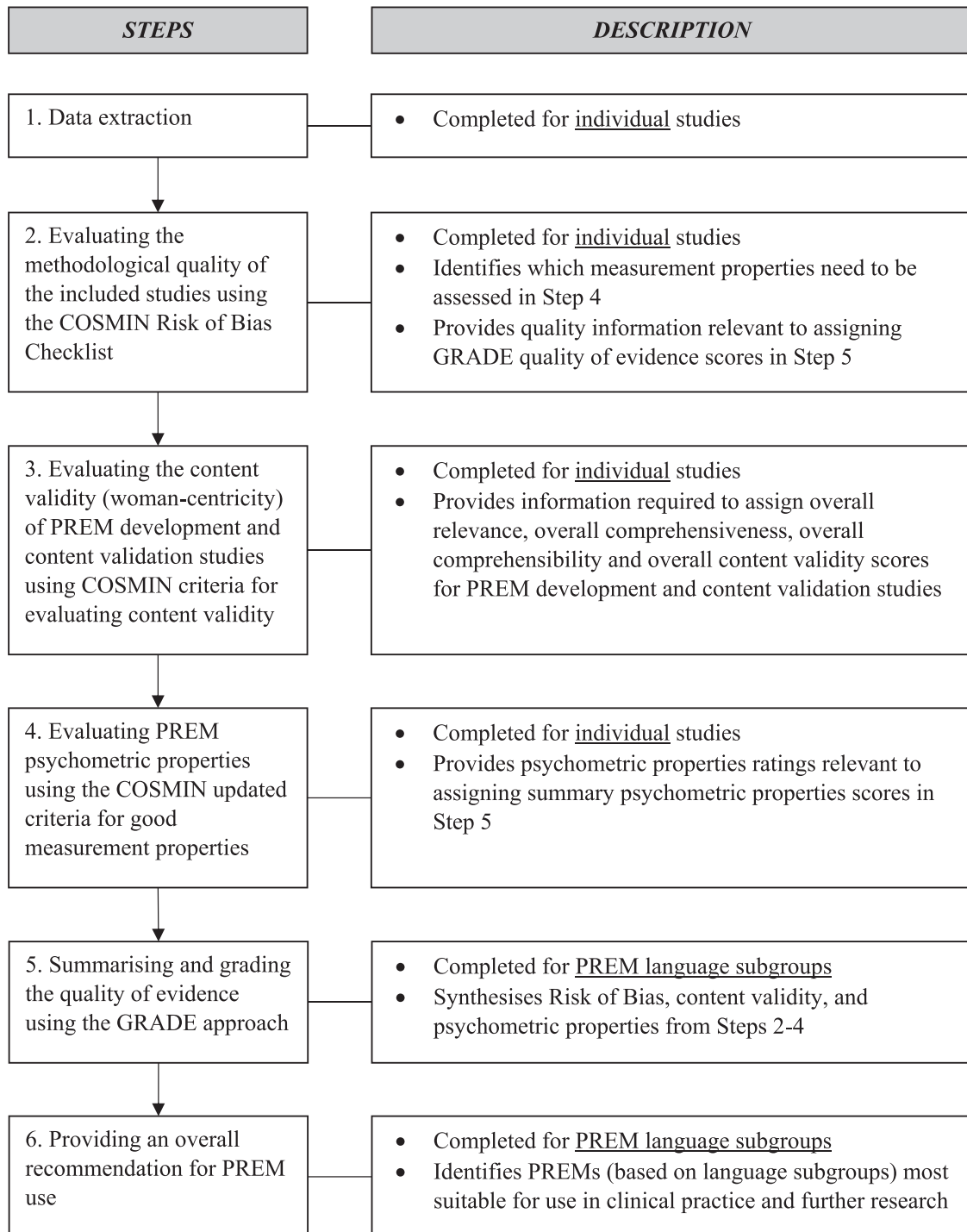
Evaluating the woman-centricity (content validity) of PREM development and content validation studies

Content validity is the most important measurement property of a PREM. It refers to whether the content of an instrument is an adequate reflection of the phenomenon being measured in the

eyes of the instrument user.¹⁶ In the case of maternity PREMs, instrument users are women, emphasizing the importance of women-centricity in PREM development and content validation studies. Each PREM development and content validation study was evaluated for relevance (5 items), comprehensiveness (1 item), and comprehensibility (4 items) to childbearing women using the COSMIN criteria for evaluating content validity.¹⁶ Each item is rated as having sufficient (+), insufficient (−), or indeterminate evidence (?) according to the following criteria:

1. Relevance criteria
 - Are the included items relevant to maternity care?
 - Are the included items relevant to childbearing women?
 - Are the response options appropriate? (ie, a justification is given for the response options used)
 - Is the recall period appropriate? (ie, a justification is given for the duration of the recall period)
2. Comprehensiveness criteria
 - Are all key concepts included?
3. Comprehensibility criteria
 - Are the instrument instructions understood by childbearing women as intended?
 - Are the items and response options understood by childbearing women as intended?

FIGURE 1
Key steps undertaken in the conduct of this systematic review



COSMIN, Consensus-based Standards for the selection of health Measurement INstruments; *GRADE*, Grading of Recommendations Assessment, Development and Evaluation; *PREM*, patient-reported experience measure.

Bull. Systematic review of maternity patient-reported experience measures. Am J Obstet Gynecol MFM 2023.

- Are the items appropriately worded? (as judged by the review team)
- Do the response options match the question? (as judged by the review team)

Each study was then given an overall relevance, comprehensiveness, and comprehensibility score of sufficient (+), insufficient (−), inconsistent (±), or indeterminate evidence (?). These scores were then aggregated to produce an overall content validity score for each PREM language subgroup. A description of how these scores were assigned is presented in [Appendix 2](#).

Evaluating psychometric properties of PREMs

The psychometric properties (validity and reliability) of each study were assessed using the COSMIN updated criteria for good measurement properties.¹⁵ Structural validity, internal consistency, reliability, measurement error, cross-cultural validity/measurement invariance, criterion validity, and responsiveness were assessed against specific criteria (described in the protocol article⁹) and given a rating of sufficient evidence (+), insufficient evidence (−), or indeterminate evidence (?).

Summarizing and grading the quality of evidence

Where possible, COSMIN suggests summarizing and grading the quality of evidence using the GRADE (Grading of Recommendations Assessment, Development and Evaluation) approach.¹⁵ Whereas the previous steps focused on the discrete measurement properties, this step focused on the quality of a PREM as a whole. Because there were numerous language versions of the included PREMs, we summarized and graded the quality of evidence for PREMs in language subgroups. For example, there were 15 studies pertaining to the Childbirth Experience Questionnaire (CEQ), of which there were 11 different language versions.

Accordingly, we summarized and graded the quality of evidence for the 11 CEQ language subgroups.

For each PREM language subgroup, we assigned an overall rating for content validity, an overall rating for each of the relevant psychometric properties, and an accompanying level of evidence indicating how confident we are in the rating provided. Overall ratings for the quality of evidence for PREM language subgroups were based on GRADE levels of evidence.¹⁵ A “high” level of evidence suggests high confidence in the results; “moderate” suggests moderate confidence; “low” suggests limited confidence; and “very low” suggests very little confidence.¹⁵ All ratings are presented in [Appendix 2](#).

Formulating recommendations for PREMs

The final step required formulating recommendations for the most suitable PREM(s) for use in clinical practice, health service and system performance measurement, and further research. PREMs categorized as “A” are recommended for use because they have sufficient (+) content validity of any level of evidence and at least low-quality evidence for sufficient (+) internal consistency.¹⁵ PREMs categorized as “B” have the potential to be recommended, but require further quality assessment; they do not meet “A” or “C” recommendation criteria.¹⁵ PREMs categorized as “C” are not recommended for use because they have high-quality evidence of at least 1 insufficient (−) measurement property.¹⁵

Results

Study selection

A total of 54 studies were included in this review, describing the development and psychometric evaluation of 25 maternity PREMs (45 PREM language subgroups). A total of 36 studies were retrieved through electronic database searching and an additional 18 were identified through reference list checking ([Figure 2](#)).

Study characteristics

[Table 2](#) describes the characteristics of the included PREMs. [Appendix 2](#) details the studies underpinning the development and psychometric evaluation of included PREMs.

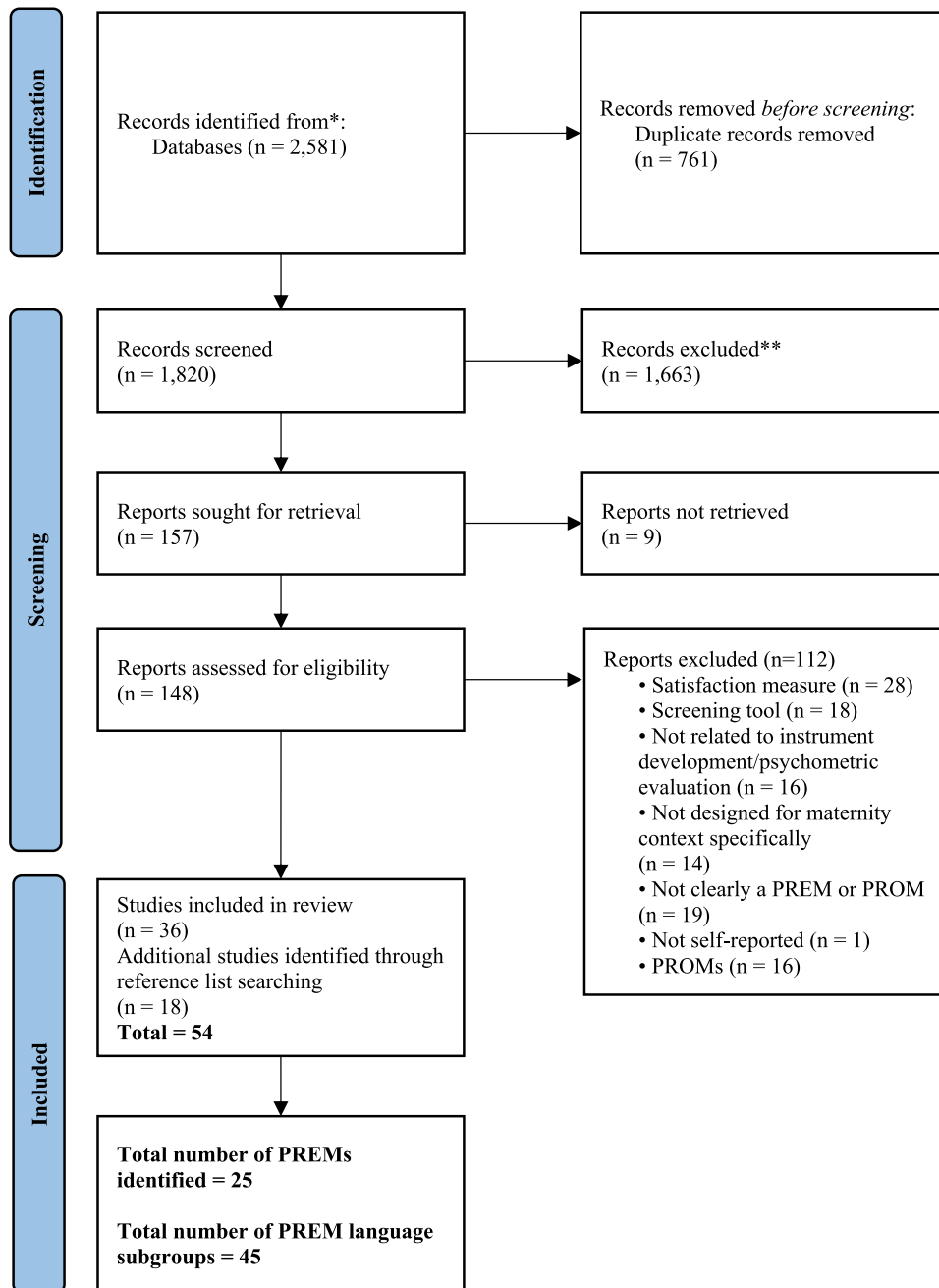
Risk of bias of included studies

Risk of bias scores for each study are tabulated in [Appendix 2](#). Scores ranged from “inadequate” to “very good.” The CEQ-English 1.0,²⁰ CEQ-English 2.0,²¹ CEQ-Spanish,²² Mother’s Autonomy in Decision Making (MADM)-Dutch,³⁷ and Mothers on Respect Index (MORI)-Dutch³⁷ were the only PREMs that received “adequate” or “very good” risk of bias scores across all of their assessments. However, they were all missing cross-cultural validity assessments. The risk of bias scores for the PREM development and content validity studies ranged from “inadequate” (lowest score) to “doubtful” (second to lowest score), indicating poor quality and high risk of bias.^{23–31,35,36,38–41,44–46,50,51,53–66,68–70,72,73} The exception to this was the Quality of Prenatal Care Questionnaire (QPCQ)-English,⁶⁷ which scored “adequate” on PREM development but “inadequate” on content validity. No studies tested PREM responsiveness or cross-cultural validity.

Synthesis of results

[Table 3](#) presents the overall woman-centricity (content validity) evidence for each PREM according to language subgroup. Two (4.4%) PREMs—the Support and Control in Birth (SCIB)-Turkish⁵⁴ and National Maternity Experience Survey (NMES)-English⁶⁴—demonstrated sufficient overall content validity, evidencing that women were appropriately involved in deciding what was relevant, comprehensive, and comprehensible to measure. However, both were underpinned by very low-quality evidence, suggesting very low confidence in the content validity result. Nine (20%) of the PREMs reported no evidence of content validity^{20–22,32,33,37,42,43,47,48,52}; 6 (13.3%) reported insufficient content validity^{46,49,58,59,61,73};

FIGURE 2
PRISMA flow diagram



Asterisk denotes MEDLINE, CINAHL Plus, PsycINFO, and Embase.

PREM, patient-reported experience measure; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; PROM, patient-reported outcome measure.

Bull. Systematic review of maternity patient-reported experience measures. Am J Obstet Gynecol MFM 2023.

and 28 (62.2%) reported inconsistent content validity evidence.^{23–31,34–36,38–41,44,45,50,51,53,55,56,60,62,63,65–72} None of the content validity results were underpinned by high-quality evidence.

Table 4 shows the overall psychometric properties evidence for each PREM

according to language subgroup, and a recommendation for use. Internal consistency reliability, structural validity, and hypothesis testing (for construct validity) were the most frequently evaluated PREM psychometric properties, undertaken in 43 (95.6%), 39 (86.7%),

and 27 (60%) of PREM language subgroups, respectively. None of the 21 (46.7%) translated PREMs underwent cross-cultural validation. Seventeen (39.5%) of 43 PREMs assessing internal consistency reliability reported a sufficient result. Twenty (51.3%) of 39

TABLE 2
Characteristics of the included maternity patient-reported experience measures (n=25)

Maternity PREM	PREM subscales (item numbers ^a)	Aspect of maternity care continuum examined	Recall period	Target population	Total number of items/score range ^a	Response options ^a	Available languages
Childbirth Experience Questionnaire (CEQ) ^{20–34}	1. Own capacity (7) 2. Professional support (4) 3. Perceived safety (6) 4. Participation (3)	Labor and childbirth	Between 24 h and 6 mo postpartum	Primiparous and multiparous women who recently labored and gave birth to a live infant with or without instruments and/or augmentation	Total=22 Subscale scores are averaged (no score range); higher scores indicate more positive experiences ^b	4-point Likert scale, ranging between 1 (totally agree) and 4 (totally disagree)	Brazilian, Chinese, Danish, English, Farsi, Malay, Mongolian, Portuguese, Sinhala, Spanish, Swedish, Turkish
Experience of Maternity Care (EMC) Survey ³⁵	1. Pregnancy scale (12) 2. Labor and birth scale (12) 3. Postnatal scale (12)	Entire maternity care continuum	Between 3 and 6 mo postpartum	Primiparous and multiparous women who recently gave birth to a live infant in hospital, a midwifery-led unit or birth center	Total=36 Scores range between 0 and 48 for each subscale; higher scores indicate more positive experiences ^b	5-point Likert scale, 1 (strongly agree) to 5 (strongly disagree)	English
Iranian Women's Childbirth Experience Questionnaire (IWCBEQ) ³⁶	1. Professional support (15) 2. Preparation (7) 3. Infant (5) 4. Positive perception (10) 5. Husband and other important support (5) 6. Control (7) 7. Fear (3)	Childbirth	Between 12 h and 2 mo postpartum	Primiparous and multiparous women with uncomplicated vaginal delivery	Total=52 Subscale and total scores are averaged (no score range); higher scores indicate more positive experiences	5-point Likert scale, 1 (totally disagree) to 5 (totally agree)	Farsi ^c
Mother's Autonomy in Decision Making (MADM) and Mothers On Respect index (MORi) ^{37–40}	MADM 1. Women's experiences of decision-making during maternity care (7) MORi 1. Women's experiences of respect and self-determination in maternity care (14)	Entire maternity care continuum	Between 1 mo and 10 y postpartum	Primiparous and multiparous women receiving care across primary care and hospital settings, and varying models of maternity care with different care providers	MADM total=7 MORi total=14 Total MADM scores range between 7 and 42, where scores between 7 and 15=very low autonomy; 16 and 24=low autonomy; 25 and 33=moderate autonomy; and 34 and 42=high autonomy Total MORi scores range between 14 and 84, where scores between 14 and 31=very low respect; 32 and 49=low respect; 50 and 66=moderate respect; and 67 and 84=high respect	Both use a 6-point Likert scale, 1 (strongly disagree) to 6 (strongly agree)	Dutch, English

(continued)

TABLE 2
Characteristics of the included maternity patient-reported experience measures (n=25) (continued)

Maternity PREM	PREM subscales (item numbers ^a)	Aspect of maternity care continuum examined	Recall period	Target population	Total number of items/score range ^a	Response options ^a	Available languages
Person-Centered Maternity Care (PCMC) Scale ^{41–43}	1. Dignity and respect (6) 2. Communication and autonomy (9) 3. Supportive care (12)	Childbirth	Between immediately postpartum and 9 wk postpartum	Primiparous and multiparous women giving birth in rural and urban health facilities, at home, or government facilities	Total=27 (short version=13) Total PCMC score ranged between 0 and 81, where subscale scores range between 0 and 18 (dignity and respect); 0 and 27 (communication and autonomy); and 0 and 36 (supportive care); higher scores indicate more positive experiences ^b	4-point Likert scale, 1 (no, never) to 4 (yes, all of the time), and a “not applicable” option for some items	Hindi, Luo, Swahili
Person-Centered Prenatal Care (PCPC) Scale ⁴⁴	1. Dignity and respect (10) 2. Communication and autonomy (14) 3. Responsive and supportive care (10)	Antenatal	Up to 1 y postpartum	Predominantly Black primiparous and multiparous women who received antenatal care in a hospital-based or community clinic	Total=34 Total PCPC score ranges between 0 and 100 (standardized); higher scores indicate better experiences ^b	4-point Likert scale, 0 (no, never) to 3 (yes, all the time)	English
Questionnaire for Assessing the Childbirth Experience (QACE) - Short version ^{45–48}	1. Intrapartum care (7) 2. First moments with the newborn (3) 3. Positive childbirth feelings (5) 4. Negative childbirth feelings (5)	Childbirth	Between 1 and 3 mo postpartum	Primiparous and multiparous women giving birth to a live infant via vaginal, instrumental, cesarean delivery, and water birth	Total=20 Subscale scores are averaged (no score range); higher scores indicate more positive experiences ^b	4-point Likert scale, 1 (totally) to 4 (not at all)	Farsi, French, Spanish
Quality of Respectful Maternity Care Questionnaire in Iran (QRMCI) ⁴⁹	1. Labor (41) 2. Delivery (10) 3. Postpartum (8)	Labor, childbirth, and postpartum	Up to 8 wk postpartum	Women referred to hospital-affiliated healthcare centers for after-care services of delivery	Total=59 No scoring scheme described	5-point Likert scale, 0 (never) to 4 (all of the time)	Farsi ^c
Respectful Maternity Care (RMC) Scale ^{50–52}	1. Abusive care (4) 2. Effective care (4) 3. Friendly care (3) 4. Respectful communication (3)	Entire maternity care continuum	Between 6 and 7 wk postpartum	Primiparous and multiparous women giving birth to a live infant via cesarean delivery and vaginal birth, admitted to postpartum wards across urban-based public and private facilities	Total=14 No scoring scheme described	5-point Likert scale, 1 (strongly disagree) to 5 (strongly agree)	English, Farsi ^c , Persian

(continued)

TABLE 2

Characteristics of the included maternity patient-reported experience measures (n=25) (continued)

Maternity PREM	PREM subscales (item numbers ^a)	Aspect of maternity care continuum examined	Recall period	Target population	Total number of items/score range ^a	Response options ^a	Available languages
Support and Control in Birth (SCIB) Questionnaire ^{53–55}	1. Internal control (10) 2. External control (11) 3. Support (12)	Childbirth	Between 24 h and 3 y postpartum	Primiparous and multiparous women who did or did not experience complications during pregnancy, labor, and childbirth, and gave birth in a hospital facility	Total=33 Subscale scores are averaged (no score range); higher scores indicate more positive experiences ^b	5-point Likert scale, 1 (strongly disagree) to 5 (strongly agree)	English, Mandarin Chinese, Turkish
Women's Experience in Childbirth Survey (WECS) ⁵⁶	1. Support during childbirth (16) 2. Physical and emotional responses to childbirth (11) 3. Transformative childbirth experience (9) 4. Ability to handle pain (6)	Childbirth	Between immediately and 4 wk postpartum	Primiparous and multiparous women who had a full-term (≥ 37 wk) pregnancy resulting in a live, singleton, vaginal birth	Total=42 Total WECS score ranges between 49 and 245; higher scores indicate more positive experiences ^b	5-point Likert scale, 1 (strongly disagree) to 5 (strongly agree)	English
Childbirth Perception Scale (CPS) ⁵⁷	1. Perception of delivery (6) 2. Perception of the first postpartum week (6)	Childbirth and early postpartum	Up to 7 d postpartum	Low-risk White primiparous and multiparous women who gave birth at home and in hospital via spontaneous, induced, vacuum-assisted, forceps-assisted, primary or secondary cesarean delivery	Total=12 No scoring scheme described; higher scores indicate poorer experiences	4-point Likert scale, 0 (totally agree) to 3 (totally disagree)	Dutch
Delivery Perception Questionnaire (DPQ) ⁵⁸	1. Obstetrical support (1) 2. Perceptions experienced during labor and delivery (8) 3. Emotions experienced during labor and delivery (4) 4. Worry about women's self (1) 5. Catastrophic thoughts about the infant (2) 6. Overall delivery perception (2)	Labor and childbirth	Between 24 and 48 h postpartum	Low-risk, nulliparous women whose delivery occurred at 37 weeks' gestation and were admitted to the postpartum ward of a university hospital	Total=18 (indices) No scoring scheme described; higher scores indicate more positive experiences ^b	4-point Likert scale, 1 (nothing at all) to 4 (completely)	Italian

(continued)

TABLE 2
Characteristics of the included maternity patient-reported experience measures (n=25) (continued)

Maternity PREM	PREM subscales (item numbers ^a)	Aspect of maternity care continuum examined	Recall period	Target population	Total number of items/score range ^a	Response options ^a	Available languages
Early Labour Experiences Questionnaire (ELEQ) ^{59,60}	Primiparous version 1. Emotional well-being (7) 2. Emotional distress (6) 3. Perceptions of nursing care (10) Multiparous version 1. Emotional well-being (6) 2. Emotional distress (6) 3. Perceptions of nursing care (8) 4. Teamwork (2)	Labor	Between 2 and 3 mo postpartum	Primiparous and multiparous women who had a spontaneous labour onset after 37 weeks' gestation	Primiparous version total=23 Multiparous version total=22 Items scored individually with averages; higher scores indicate more positive experiences ^b	5-point Likert scale, 1 (yes, definitely) to 5 (not at all), and a "not applicable" option	English, Swedish
Kolling Survey ⁶¹	1. Antenatal (?) 2. Birth (?) 3. Postnatal (?)	Entire maternity care continuum	Up to 2 y postpartum	Primiparous and multiparous women representative of all women giving birth across 7 public maternity units in 2 local health districts	Total=123 Total Kolling Survey score ranges between 0 and 100 (recoded); higher scores indicate optimal performance	Not reported	English
Mother-Centered Pregnancy Care (MCPC) Survey ⁶²	1. Screening (5) 2. Prenatal care (16) 3. Birth care (35) 4. Postpartum care (12) 5. Care preferences (22) 6. Informal payments (11) 7. Mother autonomy in decision-making (8)	Entire maternity care continuum	Not reported	Primiparous and multiparous women with children aged <5 y	Total=111 No scoring scheme described	Not reported	Hungarian
Measurement of Midwifery Quality Postpartum (MMAY-postpartum) ⁶³	1. Personal control (3) 2. Trusting relationship (7) 3. Orientation and security (6)	Postpartum	Between 4 and 12 mo postpartum	Primiparous and multiparous women from a random selection of obstetrical departments	Total=16 Total MMAY postpartum score ranges between 16 and 80, where subscale scores range between 7 and 35 (trusting relationship); 6 and 30 (orientation and security); and 3 and 15 (personal control); higher scores indicate more positive experiences	5-point Likert scale, 1 (not applicable) to 5 (fully applicable)	German ^c

(continued)

TABLE 2

Characteristics of the included maternity patient-reported experience measures (n=25) (continued)

Maternity PREM	PREM subscales (item numbers ^a)	Aspect of maternity care continuum examined	Recall period	Target population	Total number of items/score range ^a	Response options ^a	Available languages
National Maternity Experience Survey (NMES) Ireland ⁶⁴ NOTE: this was a concept elicitation study where the NMES was developed but not administered	1. Care during pregnancy (?) 2. Care during labor and birth (?) 3. Care in hospital after the birth of infant (?) 4. Specialized care for infant (?) 5. Feeding the infant (?) 6. Care at home after the birth of infant (?) 7. Overall care (?) 8. You and your household (?)	Entire maternity care continuum	Between 3 and 4 mo postpartum	Not reported	Total=95 No scoring scheme described	Not reported	English
Pregnancy and Childbirth Questionnaire (PCQ) ⁶⁵	1. Pregnancy: personal treatment (11) 2. Pregnancy: educational information (7) 3. Delivery: personal treatment (7)	Antenatal, labor, childbirth	Up to 6 wk postpartum	Primiparous and multiparous women recently giving birth to a live infant in an obstetrical department of a hospital, who gave birth via spontaneous delivery, induced labour, assisted delivery (forceps/vacuum) or cesarean delivery	Total=25 Total PCQ score ranges between 63 and 125, where subscale scores range between 24 and 55 (pregnancy – personal treatment); 8 and 35 (pregnancy – educational information); and 12 and 35 (delivery – personal treatment); higher scores indicate higher quality of care	5-point Likert scale, 1 (totally agree) to 5 (totally disagree)	Dutch
Pregnancy and Maternity care Patients' Experience Questionnaire (PreMaPEQ) ⁶⁶	1. Pregnancy control (18) 2. Birth (12) 3. Postnatal hospital stay (17) 4. Public health clinic (12)	Entire maternity care continuum	After 17 wk postpartum	Primiparous and multiparous women who gave birth across a range of Norwegian institutions via vaginal, emergency cesarean, and planned cesarean delivery	Total=59 Subscale range between 0 and 100 (linear transformation); higher scores indicate more positive experiences	5-point Likert scale, 1 (not at all) to 5 (to a very large extent), and a “not applicable” option	English

(continued)

TABLE 2
Characteristics of the included maternity patient-reported experience measures (n=25) (continued)

Maternity PREM	PREM subscales (item numbers ^a)	Aspect of maternity care continuum examined	Recall period	Target population	Total number of items/score range ^a	Response options ^a	Available languages
Quality of Prenatal Care Questionnaire (QPCQ) ⁶⁷	1. Information sharing (9) 2. Anticipatory guidance (11) 3. Sufficient time (5) 4. Approachability (4) 5. Availability (5) 6. Support and respect (12)	Antenatal	Up to 6 wk postpartum	Primiparous and multiparous women giving birth to a live infant with at least 3 prenatal care visits in a private office, clinic, or outpatient department of a hospital	Total=46 Total QPCQ and subscale scores are averaged (no score range); higher scores indicate a higher rating of care ^b	5-point Likert scale, 1 (strongly disagree) to 5 (strongly agree)	English
Intrapartal-specific Quality from the Patient's Perspective (QPP-I) ⁶⁸	1. Medical-technical competence (4) 2. Physical-technical conditions (5) 3. Identity-oriented approach (19) 4. Sociocultural atmosphere (4)	Labor and childbirth	Up to 2 mo postpartum	Primiparous and multiparous women giving birth to a live infant via vaginal, forceps, planned cesarean, and emergency cesarean delivery	Total=32 Total QPP-I scores (whether as "perceived reality" or "subjective importance") range between 0 and 10; higher scores indicate higher perceived reality and subjective importance, respectively	Two response formats for all items based on whether items are asked about "perceived quality of care" (4-point Likert scale, 1 [do not agree at all] to 4 [completely agree]) or "subjective importance ascribed by women" (4-point Likert scale, 1 [of little or no importance] to 4 [of the very highest importance]); both used a "not applicable" option	Swedish
Responsiveness in Perinatal and Obstetric Health Care Questionnaire (ReproQ) ⁶⁹⁻⁷¹	1. Dignity (?) 2. Autonomy (?) 3. Confidentiality (?) 4. Communication (?) 5. Prompt attention (?) 6. Social consideration (?) 7. Quality of basic amenities (?) 8. Choice and continuity (?)	Entire maternity care continuum	Antenatal version: 34-weeks' gestation Postpartum version: between 2 and 8 wk postpartum	Currently and recently pregnant primiparous and multiparous women recruited across midwifery clinics, perinatal units (hospitals with nearby midwife practices), and postnatal care organizations	Total=33 Total ReproQ and subscale scores are averaged (no score range) or presented as the proportion of women with negative experiences (percentage negative); higher scores indicate more positive experiences	4-point Likert scale, 1 (never) to 4 (always), and a "not applicable" option	Dutch

(continued)

TABLE 2

Characteristics of the included maternity patient-reported experience measures (n=25) (continued)

Maternity PREM	PREM subscales (item numbers ^a)	Aspect of maternity care continuum examined	Recall period	Target population	Total number of items/score range ^a	Response options ^a	Available languages
Women's Perception-Respectful Maternity Care (WP-RMC) ⁷²	1. Providing comfort (7) 2. Participatory care (7) 3. Mistreatment (5)	Labor and childbirth	Not reported	Primiparous and multiparous women who had a low-risk pregnancy, normal vaginal childbirth, and gave birth to a healthy infant with normal birthweight (2500–4000 g), recruited across urban primary healthcare centers, and public and semipublic hospitals	Total=19 No scoring scheme described	Not reported	Farsi ^c
Scale of Women's perception for Supportive Care Given During Labor ⁷³	1. Comforting behaviors (15) 2. Education (8) 3. Disturbing behaviors (10)	Labor	Not reported	Primiparous and multiparous women who had a vaginal birth, and had no labor complications or complications in the first 24 hours after birth, recruited from 2 postnatal hospital clinics	Total=33 Total scale score ranged between 33 and 132, where subscale scores range between 15 and 60 (comforting behaviors); 8 and 32 (education); and 10 and 40 (disturbing behaviors); higher scores indicate more supportive nursing care ^b	4-point Likert scale, 1 (never) to 4 (always)	Turkish

(?)=subscale item numbers not reported.

PREM, patient-reported experience measure.

^a Based on most recently published study; ^b Some items require reverse scoring; ^c Assumed language as actual language because it is not reported.

Bull. Systematic review of maternity patient-reported experience measures. *Am J Obstet Gynecol MFM* 2023.

TABLE 3

Overall woman-centricity (content validity) evaluation for the included patient-reported experience measures (based on language subgroup)

PREM-language	Risk of Bias content validity evidence		Overall relevance result	Overall comprehensiveness result	Overall comprehensibility result	Overall content validity result	Overall QoE
	Quality of PREM development	Quality of content validation studies					
Childbirth Experience Questionnaire (CEQ)							
CEQ-Swedish ^b		NR	±	+	±	±	Very low
CEQ-English ^b							NE
CEQ-Spanish							NE
CEQ-Malay							NE
CEQ-Farsi ^b		Inadequate	±	-	±	±	Very low
CEQ-Turkish	Inadequate ^a	Inadequate	±	-	±	±	Very low
CEQ-Chinese		Doubtful	±	+	±	±	Moderate
CEQ-Danish		Inadequate	±	+	±	±	Very low
CEQ-Brazilian Portuguese							NE
CEQ-Sinhala		Inadequate	±	-	±	±	Very low
CEQ-Mongolian		Inadequate	±	-	±	±	Very low
Experience of Maternity Care (EMC) Survey							
EMC-English	Doubtful	NR	±	+	±	±	Low
Iranian Women's Childbirth Experience Questionnaire (IWCBEQ)							
IWCBEQ-Farsi	Inadequate	Inadequate	±	+	±	±	Very low
Mother's Autonomy in Decision Making (MADM) and Mothers On Respect index (MORI)							
MADM-English and MORI-English ^b	Inadequate ^a	Doubtful	±	-	±	±	Low
MADM-Dutch and MORI-Dutch							NE
Person-Centred Maternity Care (PCMC) Scale							
PCMC-Swahili and Luo		Inadequate	±	+	±	±	Low
PCMC Scale-Hindi	Doubtful ^a						NE
PCMC-Short-Hindi, Swahili and Luo							NE

(continued)

Person-Centred Prenatal Care (PCPC) Scale							
PCPC-English	Doubtful	Inadequate	±	+	±	±	Low
Questionnaire for Assessing the Childbirth Experience (QACE) – Short version							
QACE-Short-French		NR	?	+	±	±	Low
QACE-Short-Farsi	Doubtful ^a	Inadequate	-	-	±	-	Low
QACE-Short-Spanish ^b							NE
Quality of Respectful Maternity Care Questionnaire in Iran (QRMCI)							
QRMCI-Farsi	Inadequate	Inadequate	-	-	±	-	Very low
Respectful Maternity Care (RMC) Scale							
RMC-English		Inadequate	±	-	±	±	Very low
RMC-Farsi	Inadequate ^a	Inadequate	±	-	±	±	Very low
RMC-Persian							NE
Support and Control in Birth (SCIB) Questionnaire							
SCIB-English		NR	±	-	±	±	Very low
SCIB-Turkish	Inadequate ^a	Inadequate	+	+	±	+	Very low
SCIB-Chinese		Inadequate	+	-	±	±	Very low
Women's Experience of Childbirth Survey (WECS)							
WECS-English	Inadequate	Doubtful	±	+	±	±	Moderate
Childbirth Perception Scale (CPS)							
CPS-Dutch	Inadequate	NR	+	-	±	±	Very low
Delivery Perception Scale							
DPQ-Italian	Inadequate	NR	-	-	±	-	Very low
Early Labour Experiences Questionnaire (ELEQ)							
ELEQ-English		Inadequate	-	-	±	-	Low
SWE(Swedish)-ELEQ-PP and SWE-ELEQ-MP	Doubtful ^a	Inadequate	±	-	±	±	Low
Kolling Survey							
Kolling-English	Inadequate	NR	-	-	±	-	Very low
Mother-Centred Pregnancy Care (MPCP) Survey							
MCPC-Hungarian	Inadequate	NR	±	-	±	±	Very low
Measurement of Midwifery Quality Postpartum (MMAY-Postpartum)							
MMAY-Postpartum-German	Inadequate	NR	±	-	±	±	Very low

(continued)

National Maternity Experience Survey (NMES) Ireland							
NMES-English	Inadequate	Inadequate	+	+	±	+	Very low
Pregnancy and Childbirth Questionnaire (PCQ)							
PCQ-Dutch	Doubtful	NR	±	-	±	±	Low
Pregnancy and Maternity care Patients' Experience Questionnaire (PreMaPEQ)							
PreMaPEQ-English	Inadequate	NR	+	-	±	±	Very low
Quality of Prenatal Care Questionnaire (QPCQ)							
QPCP-English	Adequate	Inadequate	±	+	±	±	Moderate
Intrapartal-specific Quality from the Patient's Perspective (QPP-I)							
QPP-I-Swedish	Inadequate	NR	±	-	±	±	Very low
Responsiveness in Perinatal and Obstetric Health Care Questionnaire (ReproQ)							
ReproQ-Dutch ^b	Inadequate	Doubtful	±	-	±	±	Moderate
Women's Perception-Respectful Maternity Care (WP-RMC)							
WP-RMC-Farsi	Inadequate	Inadequate	±	+	±	±	Very low
Scale of Women's perceptions for Supportive Care Given During Labour							
Scale of Women's perceptions for Supportive Care Given During Labour-Turkish	Inadequate	Inadequate	-	-	±	-	Very low
Indicates that a content validity quality was not reported							
*Result' refers to overall performance on the specified aspect of content validity as either: sufficient (+), insufficient (-), inconsistent (±) or indeterminate (?)							
*QoE' refers to the quality of evidence using GRADE, reported as: High, Moderate, Low or Very Low							
^a PREM development risk of Bias only assessed in first study (all other versions are a derivative of the original PREM version)							
^b Multiple versions of the same PREM language version informed this assessment (please refer to Appendix 2)							
PREM = Patient-Reported Experience Measure; NR = Not reported; NE = No evidence							
Bull. Systematic review of maternity patient-reported experience measures. Am J Obstet Gynecol MFM 2023.							

PREMs assessing structural validity reported a sufficient result. Six (22.2%) of 27 PREMs undergoing hypothesis testing (for construct validity) reported a sufficient result. Only the Women's Experience in Childbirth Survey (WECS)-English demonstrated sufficient test-retest reliability⁵⁶; all other instruments assessing test-retest reliability used inappropriate statistical approaches for the type of response data available.^{20,21,24-27,29,30,33,46-48,51,52,54,55,66,67,69-72} No PREMs underwent evaluation for responsiveness.

The SCIB-Turkish⁵⁴ is the only PREM with an "A" recommendation level. The SCIB-Turkish demonstrated sufficient content validity (of very low-quality evidence), sufficient

structural validity (of low-quality evidence), and sufficient internal consistency reliability (of high-quality evidence). Over 75% (34) of PREM language subgroups received a level "B" recommendation. Of these, 15 (44.1%) received a level "B" recommendation due to insufficient (-), inconsistent (±), or no evidence of content validity. Ten (22.2%) PREM language subgroups are not recommended for use (level "C").^{20,21,24-27,31,35,37,55-57,68}

Comment

Main findings

Measuring what matters to women during their maternity care experiences is crucial to supporting robust, woman-

centered health service and system performance measurement. Moreover, there is growing evidence to illustrate how PREMs can support real-world clinical care through open dialogue and decision-making that aligns with what matters most to women.¹³ This study sought to identify and critically appraise the risk of bias, woman-centricity (content validity), and psychometric properties of maternity PREMs published in the scientific literature. Overall, we identified 54 studies reporting on the development and psychometric evaluation of 25 PREMs. To enable fair and consistent result comparisons, PREMs were grouped into language subgroups, of which there were 45. Only the SCIB-Turkish⁵⁴ demonstrated sufficient

TABLE 4

Overall psychometric properties evaluation for the included patient-reported experience measures (based on language subgroup) and overall recommendation for use

PREM-language	Structural Validity		Internal consistency		Cross-cultural validity		Reliability		Measurement error		Criterion validity		Hypothesis testing		Responsiveness		Recommendation
	Result	QoE	Result	QoE	Result	QoE	Resu	Qo	Result	QoE	Result	QoE	Resu	Qo	Result	QoE	
							It	E					It	E			
Childbirth Experience Questionnaire (CEQ)																	
CEQ-Swedish ^a	+	H	+	H	NA		NE		NE		NE		+	M	NE		B
CEQ-English ^a	NE		?	H	NE		-	H	NE		?	H	?	H	NE		C
CEQ-Spanish	+	H	?	H	NE		NE		NE		NE		?	H	NE		B
CEQ-Malay	?	L	-	VL	NE		NE		NE		NE		NE		NE		B
CEQ-Farsi ^b	+	M	?	H	NE		-	VL	NE		NE		-	H	NE		C
CEQ-Turkish	+	H	-	H	NE		-	L	NE		NE		NE		NE		C
CEQ-Chinese	+	L	-	VL	NE		NE		NE		NE		?	L	NE		B
CEQ-Danish	+	M	+	H	NE		?	M	NE		NE		?	H	NE		B
CEQ-Brazilian Portuguese	NE		?	H	NE		?	VL	?	VL	NE		?	H	NE		B
CEQ-Sinhala	?	VL	?	H	NE		-	M	NE		NE		?	L	NE		B
CEQ-Mongolian	+	H	-	H	NE		NE		NE		NE		?	L	NE		C
Experience of Maternity Care (EMC) Survey																	
EMC-English	-	L	-	H	NA		NE		NE		NE		?	H	NE		C
Iranian Women's Childbirth Experience Questionnaire (IWCBEQ)																	
IWCBEQ-Farsi	?	VL	?	H	NA		NE		NE		NE		NE		NE		B
Mother's Autonomy in Decision Making (MADM) and Mothers On Respect index (MORi)																	
MADM-English and MORi-English ^a	?	L	+	H	NA		NE		NE		NE		?	L	NE		B
MADM-Dutch and MORi-Dutch	NE		+	H	NE		NE		NE		NE		-	H	NE		C
Person-Centred Maternity Care (PCMC) Scale																	
PCMC-Swahili and Luo	?	L	?	H	NE		NE		NE		NE		?	VL	NE		B
PCMC Scale-Hindi	?	H	?	H	NE		NE		NE		NE		?	VL	NE		B

(continued)

PCMC-Short-Hindi, Swahili and Luo	-	VL	?	H	NA	NE	NE	NE	NE	?	VL	NE	B
Person-Centered Prenatal Care (PCPC) Scale													
PCPC-English	+	M	+	H	NA	NE	NE	NE	NE	+	VL	NE	B
Questionnaire for Assessing the Childbirth Experience (QACE) - Short version													
QACE-Short-French	?	VL	?	H	NA	NE	NE	NE	NE	?	VL	NE	B
QACE-Short-Farsi	+	VL	+	H	NE	?	VL	NE	NE	?	VL	NE	B
QACE-Short-Spanish ^a	+	VL	-	M	NE	?	VL	NE	NE	+	VL	NE	B
Quality of Respectful Maternity Care Questionnaire in Iran (QRMCI)													
QRMCI-Farsi	-	L	?	L	NA	NE	NE	NE	NE	NE	NE	NE	B
Respectful Maternity Care (RMC) Scale													
RMC-English	+	M	+	H	NA	NE	NE	NE	NE	?	VL	NE	B
RMC-Farsi	+	L	+	L	NE	?	VL	NE	NE	NE	NE	NE	B
RMC-Persian	+	VL	-	L	NE	?	VL	NE	NE	NE	NE	NE	B
Support and Control in Birth (SCIB) Questionnaire													
SCIB-English	+	H	+	H	NA	NE	NE	NE	NE	NE	NE	NE	B
SCIB-Turkish	+	L	+	H	NE	?	VL	NE	NE	NE	NE	NE	A
SCIB-Chinese	-	M	?	H	NE	?	VL	NE	-	H	NE	NE	C
Women's Experience of Childbirth Survey (WECS)													
WECS-English	-	VL	-	H	NA	+	L	NE	NE	?	VL	NE	C
Childbirth Perception Scale (CPS)													
CPS-Dutch	-	H	?	H	NA	NE	NE	NE	NE	?	L	NE	C
Delivery Perception Questionnaire (DPQ)													
DPQ-Italian	+	L	+	H	NA	NE	NE	NE	NE	NE	NE	NE	B
Early Labour Experiences Questionnaire (ELEQ)													
ELEQ-English	+	M	+	H	NA	NE	NE	NE	NE	+	L	NE	B
SWE(Swedish)-ELEQ-PP and SWE-ELEQ-MP	+	L	?	L	NE	NE	NE	NE	NE	NE	NE	NE	B
Kolling Survey													

(continued)

Kolling-English	NE	NE	NA	NE	NE	NE	NE	NE	NE	B				
Mother-Centred Pregnancy Care (MCPC) Survey														
MCPC-Hungarian	NE	NE	NA	NE	NE	NE	NE	NE	NE	B				
Measurement of Midwifery Quality Postpartum (MMAY-postpartum)														
MMAY-Postpartum-German	+	VL	+	H	NA	NE	NE	NE	+	H	NE	B		
National Maternity Experience Survey (NMES) Ireland														
NMES-English	NE	NE	NA	NE	NE	NE	NE	NE	NE	NE	B			
Pregnancy and Childbirth Questionnaire (PCQ)														
PCQ-Dutch	?	L	+	H	NA	NE	NE	NE	?	L	NE	B		
Pregnancy and Maternity care Patients' Experience Questionnaire (PreMaPEQ)														
PreMa-PEQ-English	-	L	?	H	NA	?	VL	NE	NE	NE	NE	B		
Quality of Prenatal Care Questionnaire (QPCQ)														
QPCQ-English	?	H	+	H	NA	?	VL	NE	NE	+	H	NE	B	
Intrapartal-specific Quality from the Patient's Perspective (QPP-I)														
QPP-I-Swedish	?	H	-	H	NA	NE	NE	NE	NE	NE	NE	C		
Responsiveness in Perinatal and Obstetric Health Care Questionnaire (ReproQ)														
ReproQ-Dutch ^a	?	VL	-	VL	NE	?	VL	-	M	NE	?	M	NE	B
Women's Perception-Respectful Maternity Care (WP-RMC)														
WP-RMC-Farsi	+	VL	+	VL	NA	?	VL	NE	NE	NE	NE	B		
Scale of Women's perception for Supportive Care Given During Labour														
Scale of Women's perception for Supportive Care Given During Labour-Turkish	?	VL	+	H	NA	NE	NE	NE	NE	NE	NE	B		

'Result' refers to overall performance on the specified measurement property as either: sufficient (+), insufficient (-) or indeterminate (?)

'QoE' (Quality of Evidence) refers to the quality of evidence using GRADE, reported as: High (H), Moderate (M), Low (L) or Very Low (VL)

'Not applicable' in cross-cultural validity indicates that demonstrating this psychometric property was not necessary

'Recommendation' refers to whether a PREM is suitable for use in a real-world application, reported as: Recommended for use (A), Potential for use but requires further testing (B), Not recommended for use (C)

^aMultiple versions of the same PREM language version informed this assessment (see Appendix 2)

PREM = Patient-Reported Experience Measure; NA = Not applicable; NE = No evidence

Bull. Systematic review of maternity patient-reported experience measures. Am J Obstet Gynecol MFM 2023.

evidence to receive a level “A” recommendation. Consequently, this is the only PREM that can be recommended at this time for use in clinical practice, health service and system performance measurement, and further research. The results of other PREM language subgroups evidenced significant variation in risk of bias, content validity, psychometric properties, and quality of evidence. The following paragraphs discuss the strengths and limitations of this review followed by several recommendations for improvement in the field of maternity PREMs. In addition, we discuss the trade-offs between implementing a standardized measure that will enable the field of maternity experience measurement to advance, relative to waiting for the “right” maternity PREM to be developed.

Strengths and limitations

There were several strengths to this review. First, the search strategy was robust, complemented by reference list hand-searching. This ensured that all available evidence in the scientific literature supporting the development and psychometric evaluation of the maternity PREMs was included. Second, the triple-tier COSMIN approach to evaluating risk of bias, woman-centricity (content validity), and psychometric properties enabled us to holistically examine the quality of each study and PREM as a whole. Moreover, summarizing results for PREM language subgroups enabled us to provide readers with overall recommendations for PREM use. Finally, article screening, data extraction, and all 3 stages of quality assessment (risk of bias, content validity, and psychometric properties appraisal) involved 2 reviewers, ensuring accuracy of the results.

One limitation of this review was the “worst score counts” principle applied in the risk of bias assessment. Although this is the recommended approach by COSMIN and minimizes ambiguity in the scoring process (ie, instead of relying on the researcher’s ability and experience to consider all criteria as a whole), it also removes any opportunity to weigh the relative importance of

different criteria. Risk of bias and quality of evidence scores, therefore, may be overly negative. Another limitation of this review was its heavy reliance on COSMIN guidance. First, COSMIN was originally developed for PROMs. However, although PROMs and PREMs capture different constructs, they are developed and psychometrically evaluated in very similar ways, thereby reinforcing why COSMIN was an appropriate approach to follow. Second, the COSMIN guidance is dogmatic and, at times, inflexible. Resultantly, most of the available maternity PREMs were rejected from a psychometric standpoint. Nevertheless, we maintain that adhering to COSMIN guidance was the most appropriate, evidence-based approach for this review because it is comprehensive, has been used in similar reviews of PREMs,^{2,74,75} and promotes a high level of rigor in instrument development and evaluation.

Comparison with existing literature

The most important finding of this review was the lack of detailed content validity evidence for the included PREMs. Consequently, roughly half of all PREMs received a level “B” recommendation. Most studies did not adequately describe how they involved women when establishing what was relevant, comprehensive, and comprehensible to capture in the instrument. Although the lack of detail may be the result of journal word limits as opposed to a lack of rigor, the resultant inference is that most maternity PREMs are not woman-centric, despite the purported goal of patient-reported measures being to “measure what matters”⁷⁶ to the target population. This is also problematic because a lack of content validity has downstream effects for all other measurement properties. For example, irrelevant items may contribute to low internal consistency reliability, poorly fitting factor structures in structural validity assessment, and reduced interpretability of an instrument to the target population.¹⁶ Moreover, instruments that ask irrelevant questions or fail to ask critical questions may result in biased responses and low response rates.¹⁶ In clinical practice, this can lead to wasted time

for both women and care providers by completing PREMs that are not meaningful and important to women. This may also jeopardize the woman–care provider dyad by focusing on aspects of the maternal and childbearing experience that are not important to women.

On the basis of our findings, we propose 2 recommendations. The first is to ensure meaningful involvement of women in the development of new maternity PREMs, particularly when defining what is relevant, comprehensive, and comprehensible to measure. Readers are encouraged to consult COSMIN guidance for assessing content validity because this illustrates important design requirements for content validity studies.¹⁶ We also recommend consulting the content validation processes described by Beecher et al⁶⁴ and Inci et al,⁵⁴ who illustrated sufficient overall content validity scores for the NMES-English and SCIB-Turkish, respectively. If developing a new maternity PREM is infeasible, our second recommendation is to establish the content validity of an existing level “B” PREM in a comparable target population. Several of the identified PREMs would have received a level “A” recommendation if they had reported sufficient content validity.^{23,29,34,38–40,44,50,51,53,58,59,63,65,67,73}

The application of rigorous content validation approaches in similar childbearing populations to produce sufficient content validity results would support the use of these PREMs.

There were also several shortcomings in the evaluation of PREM psychometric properties. First, a sufficient score for test-retest reliability was reported for only 1 measure²⁹ despite being assessed in 17 PREM language subgroups. Most studies were scored “indeterminate” because they reported intraclass correlation coefficients (ICC). Although the ICC is appropriate in the context of continuous data, the included PREMs used Likert response scales, and therefore reported ordinal response data.⁷⁷ The weighted Kappa or Pearson product–moment correlation is more appropriate¹⁵ for PREMs using Likert response scales and should be applied in future studies.

Second, despite 27 PREM language subgroups undertaking hypothesis testing for construct validity, several were scored “indeterminate.” This was because authors failed to describe the direction and magnitude of correlation that they expected to observe between their PREM and known-groups (in the case of discriminant validity assessment), or other instruments capturing related constructs (in the case of convergent validity assessment). Readers are encouraged to refer to Dencker et al,²³ Redshaw et al,³⁵ and Janssen et al⁵⁹ studies for examples of sufficiently stipulated hypotheses. A similar level of detail should be applied in future research.

Third, responsiveness was not assessed in any of the included studies. This is not surprising given that reviews of PREMs in other healthcare fields have highlighted the same finding.^{2,17,78} Responsiveness refers to the ability of an instrument to detect meaningful change in the construct being measured (ie, experiences), over time.⁷⁹ Responsiveness is most frequently associated with measures of health status, where a significant change in health status has implications for changes to clinical practice.⁸⁰ Although significant changes in women’s experiences over time may also have clinical practice implications, the recall of an experience can fluctuate with the passage of time because of both intrinsic factors (eg, self-reflection) and external factors (eg, comparing experiences with family and friends).^{81,82} Responsiveness data in the context of maternity care experiences, therefore, may be biased and of little utility for demonstrating the psychometric rigor of a maternity PREM, particularly when considering the additional resources and time burden necessitated by a second round of data collection. Consequently, we do not consider responsiveness pertinent to demonstrating psychometric rigor in maternity PREMs.

Finally, despite identifying 21 PREM language subgroups, none of the translated PREMs reported on cross-cultural validation. Cross-cultural validity is the degree to which the performance of items on a translated or culturally adapted instrument is an adequate

reflection of the performance of the items in the original version of the instrument.¹⁵ This is also referred to as equivalence.⁸³ To assess this measurement property, data from at least 2 samples are required (eg, an instrument translated from English to Spanish will require data from both samples to undergo appropriate cross-cultural validation).¹⁵ Cross-cultural validation can be assessed by examining measurement invariance or nondifferential item functioning (non-DIF).⁸³ Both examine the psychometric equivalence of constructs across groups, demonstrating that the construct has the same meaning to those groups.⁸⁴ Although most studies that undertook translation of a maternity PREM into another language did undertake some form of qualitative pilot test or cognitive interview to ensure semantic equivalence (ie, items in the new and source language have the same meaning⁸³), cross-cultural validity was not performed to the COSMIN standard. Future studies should aim to demonstrate this quality in translated PREMs.

Conclusions and implications

Maternity PREMs should measure what matters to women receiving maternity care. However, our findings illustrate that this is currently not the case. This review identified 25 maternity PREMs, comprising 45 PREM language subgroups. Only 1 instrument received an “A” level recommendation—the SCIB-Turkish—suggesting that it is suitable for use in clinical practice, health service and system performance measurement, and future research. Future maternity PREM development needs to prioritize centering women in instrument development processes, particularly in decisions regarding what is relevant, comprehensive, and comprehensible to measure. Where the SCIB-Turkish is not appropriate for use in their context, readers are encouraged to content-validate relevant “B” level PREMs in comparable target populations. Given that content validity has flow-on effects for other psychometric properties, focusing on improving the woman-centricity of

maternity PREMs is likely to improve the overall quality of these measures.

However, although we need to be considerate of PREM quality, it is also critical that we promptly consider implementing a standardized maternity PREM to advance the field of measurement more broadly. The pursuit of identifying the “perfect” maternity PREM may be inhibiting our ability to integrate standardized data collection mechanisms into clinical practice and routine performance measurement. Without standardization, opportunities to build meaningful relationships with women that emphasize their values and preferences, and learn from one another about what services and models of care promote greater WCC, remain limited. Thus, it is critical that the goal of measuring what matters is considered in relation to the reasons for measuring in the first place—to optimize women’s experiences of maternity care. ■

Supplementary materials

Supplementary material associated with this article can be found in the online version at [doi:10.1016/j.ajogmf.2023.101102](https://doi.org/10.1016/j.ajogmf.2023.101102).

REFERENCES

1. Beecher C, Devane D, White M, Greene R, Dowling M. Women’s experiences of their maternity care: a principle-based concept analysis. *Women Birth* 2020;33:419–25.
2. Bull C, Byrnes J, Hettiarachchi R, Downes M. A systematic review of the validity and reliability of patient-reported experience measures. *Health Serv Res* 2019;54:1023–35.
3. U.S. Department of Health and Human Services FDA Center for Drug Evaluation and Research. U.S. Department of Health and Human Services FDA Center for Biologics Evaluation and Research, U.S. Department of Health and Human Services FDA Center for Devices and Radiological Health. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. *Health Qual Life Outcomes* 2006;4:79.
4. Doyle C, Lennox L, Bell D. A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. *BMJ Open* 2013;3:e001570.
5. Berwick DM, Nolan TW, Whittington J. The triple aim: care, health, and cost. *Health Aff (Millwood)* 2008;27:759–69.
6. Santana MJ, Manalili K, Jolley RJ, Zelinsky S, Quan H, Lu M. How to practice person-centred

care: a conceptual framework. *Health Expect* 2018;21:429–40.

7. Bull C, Teede H, Watson D, Callander EJ. Selecting and implementing patient-reported outcome and experience measures to assess health system performance. *JAMA Health Forum* 2022;3:e220326. 4.

8. Nursing and Midwifery Board of Australia. Midwife standards for practice. Canberra, ACT: Nursing and Midwifery Board of Australia; 2018.

9. Bull C, Teede H, Carrandi L, Rigney A, Cusack S, Callander E. Evaluating the development, woman-centricity and psychometric properties of maternity patient-reported outcome measures (PROMs) and patient-reported experience measures (PREMs): a systematic review protocol. *BMJ Open* 2022;12:e058952.

10. COAG Health Council. Woman-centred care: strategic directions for Australian maternity services. 2019. Available at: <https://www.health.gov.au/sites/default/files/documents/2019/11/woman-centred-care-strategic-directions-for-australian-maternity-services.pdf>. Accessed October 5, 2022.

11. Leap N. Woman-centred or women-centred care: does it matter? *Br J Midwif* 2009;17:12–6.

12. Beecher C, Greene R, O'Dwyer L, et al. Measuring women's experiences of maternity care: a systematic review of self-report survey instruments. *Women Birth* 2021;34:231–41.

13. Depla AL, Lamain-de Ruiter M, Laureij LT, et al. Patient-reported outcome and experience measures in perinatal care to guide clinical practice: prospective observational study. *J Med Internet Res* 2022;24:e37725.

14. Mokkink LB, de Vet HCW, Prinsen CAC, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Qual Life Res* 2018;27:1171–9.

15. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018;27:1147–57.

16. Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res* 2018;27:1159–70.

17. Beattie M, Murphy DJ, Atherton I, Lauder W. Instruments to measure patient experience of healthcare quality in hospitals: a systematic review. *Syst Rev* 2015;4:97.

18. Terwee CB, Jansma EP, Riphagen II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res* 2009;18:1115–23.

19. Veritas Health Innovation. Covidence systematic review software. Veritas Health Innovation. Available at: www.covidence.org. Accessed October 5, 2022.

20. Walker KF, Wilson P, Bugg GJ, Dencker A, Thornton JG. Childbirth experience questionnaire: validating its use in the United Kingdom. *BMC Pregnancy Childbirth* 2015;15:86.

21. Walker KF, Dencker A, Thornton JG. Childbirth experience questionnaire 2: validating its use in the United Kingdom. *Eur J Obstet Gynecol Reprod Biol* 2020;5:100097.

22. Soriano-Vidal FJ, Oliver-Roig A, Cabrero-García J, Congost-Maestre N, Dencker A, Richart-Martínez M. The Spanish version of the Childbirth Experience Questionnaire (CEQ-E): reliability and validity assessment. *BMC Pregnancy Childbirth* 2016;16:372.

23. Dencker A, Taft C, Bergqvist L, Lilja H, Berg M. Childbirth experience questionnaire (CEQ): development and evaluation of a multidimensional instrument. *BMC Pregnancy Childbirth* 2010;10:81.

24. Abbaspoor Z, Moghaddam-Banaem L, Ronaghi S, Dencker A. Translation and cultural adaptation of the Childbirth Experience Questionnaire (CEQ) in Iran. *Iran J Nurs Midwifery Res* 2019;24:296–300.

25. Ghanbari-Homayi S, Dencker A, Fardiazar Z, et al. Validation of the Iranian version of the childbirth experience questionnaire 2.0. *BMC Pregnancy Childbirth* 2019;19:465.

26. Kazemi S, Dencker A, Pazandeh F, Montazeri A, Sedigh-Mobarakabadi S, Hajian S. Psychometric evaluation of the Persian version of the Childbirth Experience Questionnaire (CEQ). *BioMed Res Int* 2020;2020:6879283.

27. Mamuk R, Şahin N, Dişsiz M. The Turkish Version of the Childbirth Experience Questionnaire: Reliability and Validity Assessment. 2019;15:265–71.

28. Zhu X, Wang Y, Zhou H, Qiu L, Pang R. Adaptation of the Childbirth Experience Questionnaire (CEQ) in China: a multisite cross-sectional study. *PLoS One* 2019;14:e0215373.

29. Boie S, Lauridsen HH, Glavind J, Smed MK, Ulbjerg N, Bor P. The Childbirth Experience Questionnaire (CEQ)-Validation of its use in a Danish-speaking population of new mothers stimulated with oxytocin during labour. *PLoS One* 2020;15:e0233122.

30. Patabendige M, Palihawadana TS, Herath RP, Wijesinghe PS. Childbirth Experience Questionnaire (CEQ) in the Sri Lankan setting: translation, cultural adaptation and validation into the Sinhala language. *BMC Res Notes* 2020;13:534.

31. Parchaa T, Togoobaatar G, Fukuzawa RK, et al. Translation and validation of the Mongolian version of the childbirth experience questionnaire. *J Patient Exp* 2021;8:23743735211060636.

32. Al-Kubaisi M, Radeef A. Translation and validation study of the Malaysian version of the childbirth experience questionnaire - CEQ. *J Int Dent Med Res* 2018;11:357–61.

33. da Silva Vieira RCM, Ferreira CHJ, de Carvalho Cavalli R, do Prado MLR, Beleza ACS, Driusso P. Cross-cultural adaptation and psychometric evaluation of the Brazilian Portuguese version of the childbirth experience questionnaire. *BMC Pregnancy Childbirth* 2020;20:477.

34. Dencker A, Bergqvist L, Berg M, Greenbrook JTV, Nilsson C, Lundgren I. Measuring

women's experiences of decision-making and aspects of midwifery support: a confirmatory factor analysis of the revised Childbirth Experience Questionnaire. *BMC Pregnancy Childbirth* 2020;20:199.

35. Redshaw M, Martin CR, Savage-McGlynn E, Harrison S. Women's experiences of maternity care in England: preliminary development of a standard measure. *BMC Pregnancy Childbirth* 2019;19:167.

36. Hosseini Tabaghdehi M, Keramat A, Shahhosseini Z, Kolahdozan S, Moosazadeh M, Motaghi Z. Development and psychometric properties of Iranian women childbirth experience questionnaire. *Nurs Open* 2021;8:1360–8.

37. Feijen-de Jong EI, van der Pijl M, Vedam S, Jansen DEMC, Peters LL. Measuring respect and autonomy in Dutch maternity care: applicability of two measures. *Women Birth* 2020;33:e447–54.

38. Vedam S, Stoll K, Martin K, et al. The Mother's Autonomy in Decision Making (MADM) scale: patient-led development and psychometric testing of a new instrument to evaluate experience of maternity care. *PLoS One* 2017;12:e0171804.

39. Vedam S, Stoll K, Rubashkin N, et al. The Mothers on Respect (MOR) index: measuring quality, safety, and human rights in childbirth. *SSM Popul Health* 2017;3:201–10.

40. Jenkinson B, Kearney L, Kynn M, et al. Validating a scale to measure respectful maternity care in Australia: challenges and recommendations. *Midwifery* 2021;103:103090.

41. Afulani PA, Diamond-Smith N, Golub G, Sudhinaraset M. Development of a tool to measure person-centered maternity care in developing settings: validation in a rural and urban Kenyan population. *Reprod Health* 2017;14:118.

42. Afulani PA, Diamond-Smith N, Phillips B, Singhal S, Sudhinaraset M. Validation of the person-centered maternity care scale in India. *Reprod Health* 2018;15:147.

43. Afulani PA, Feeser K, Sudhinaraset M, Aborigo R, Montagu D, Chakraborty N. Toward the development of a short multi-country person-centered maternity care scale. *Int J Gynaecol Obstet* 2019;146:80–7.

44. Afulani PA, Altman MR, Castillo E, et al. Development of the person-centered prenatal care scale for people of color. *Am J Obstet Gynecol* 2021;225. 427.e1–13.

45. Carquillat P, Vendittelli F, Perneger T, Guttier MJ. Development of a questionnaire for assessing the childbirth experience (QACE). *BMC Pregnancy Childbirth* 2017;17:279.

46. Mirghafourvand M, Jafarabadi MA, Ghanbari-Homayi S. Adaptation of short version of questionnaire for assessing the childbirth experience (QACE) to the Iranian culture. *BMC Pregnancy Childbirth* 2020;20:616.

47. Rodríguez Coll P, Casañas R, Collado Palomares A, et al. Validation and psychometric properties of the Spanish version of the questionnaire for assessing the childbirth experience (QACE). *Sex Reprod Healthc* 2021;27:100584.

48. Rodríguez Coll P, Casañas R, Collado Palomares A, Maldonado Aubián G, Vicente García M, Escuriel Peiro R. Validation and psychometric properties of the Spanish version of the questionnaire for assessing the childbirth experience (QACE). *Sex Reprod Healthc* 2021;27:100584. <https://doi.org/10.1016/j.srhc.2020.100584>.
49. Taavoni S, Goldani Z, Rostami Gooran N, Haghani H. Development and assessment of respectful maternity care questionnaire in Iran. *Int J Community Based Nurs Midwifery* 2018;6:334–49.
50. Sheferaw ED, Mengesha TZ, Wase SB. Development of a tool to measure women's perception of respectful maternity care in public health facilities. *BMC Pregnancy Childbirth* 2016;16:67.
51. Hajizadeh K, Asghari Jafarabadi M, Vaezi M, Meedy S, Mohammad-Alizadeh-Charandabi S, Mirghafourvand M. The psychometric properties of the respectful maternity care (RMC) for an Iranian population. *BMC Health Serv Res* 2020;20:894.
52. Esmkhani M, Namadian M, Noorozy A, Korte JE. Psychometric properties of a Persian version of respectful maternity care questionnaire. *BMC Pregnancy Childbirth* 2021;21:218.
53. Ford E, Ayers S, Wright DB. Measurement of maternal perceptions of support and control in birth (SCIB). *J Womens Health (Larchmt)* 2009;18:245–52.
54. Inci F, Gokce Isbir G, Tanhan F. The Turkish version of perceived support and control in birth scale. *J Psychosom Obstet Gynaecol* 2015;36:103–13.
55. Liu SY, Lu YY, Gau ML, Liu CY. Psychometric testing of the support and control in birth scale. *BMC Pregnancy Childbirth* 2020;20:293.
56. Rini EV. The development and psychometric analysis of the women's experience in childbirth survey. *J Nurs Meas* 2016;24:268–80.
57. Truijens SE, Wijnen HA, Pommer AM, Oei SG, Pop VJ. Development of the Childbirth Perception Scale (CPS): perception of delivery and the first postpartum week. *Arch Womens Ment Health* 2014;17:411–21.
58. Mannarini S, Boffo M, Bertucci V, Andrisani A, Ambrosini G. A Rasch-based dimension of delivery experience: spontaneous vs. medically assisted conception. *J Clin Nurs* 2013;22:2404–16.
59. Janssen PA, Desmarais SL. Development and psychometric properties of the Early Labour Experience Questionnaire (ELEQ). *Midwifery* 2013;29:181–9.
60. Ångeby K, Sandin-Bojö AK, Persenius M, Wilde-Larsson B. Early labour experience questionnaire: psychometric testing and women's experiences in a Swedish setting. *Midwifery* 2018;64:77–84.
61. Todd AL, Aitken CA, Boyd J, Porter M. Testing a health research instrument to develop a state-wide survey on maternity care. *Public Health Res Pract* 2016;26:e2611609.
62. Rubashkin N, Szebik I, Baji P, Szántó Z, Susánszky É, Vedam S. Assessing quality of maternity care in Hungary: expert validation and testing of the mother-centered prenatal care (MCPC) survey instrument. *Reprod Health* 2017;14:152.
63. Peters M, Kolip P, Schäfers R. A questionnaire to measure the quality of midwifery care in the postpartum period from women's point of view: development and psychometric testing of MMAYpostpartum. *BMC Pregnancy Childbirth* 2021;21:412.
64. Beecher C, Drummond L, Foley C, et al. Development of a survey instrument to evaluate women's experiences of their maternity care. *Women Birth* 2021;34:e396–405.
65. Truijens SE, Pommer AM, van Runnard Heimel PJ, Verhoeven CJ, Oei SG, Pop VJ. Development of the Pregnancy and Childbirth Questionnaire (PCQ): evaluating quality of care as perceived by women who recently gave birth. *Eur J Obstet Gynecol Reprod Biol* 2014;174:35–40.
66. Sjetne IS, Iversen HH, Kjøllesdal JG. A questionnaire to measure women's experiences with pregnancy, birth and postnatal care: instrument development and assessment following a national survey in Norway. *BMC Pregnancy Childbirth* 2015;15:182.
67. Heaman M, Sword WA, Akhtar-Danesh N, et al. Quality of prenatal care questionnaire: instrument development and testing. *BMC Pregnancy Childbirth* 2014;14:188.
68. Wilde-Larsson B, Larsson G, Kvist LJ, Sandin-Bojö AK. Women's opinions on intrapartum care: development of a theory-based questionnaire. *J Clin Nurs* 2010;19:1748–60.
69. van der Kooy J, Valentine NB, Birnie E, et al. Validity of a questionnaire measuring the world health organization concept of health system responsiveness with respect to perinatal services in the Dutch obstetric care system. *BMC Health Serv Res* 2014;14:622.
70. Scheerhagen M, van Stel HF, Birnie E, Franx A, Bonsel GJ. Measuring client experiences in maternity care under change: development of a questionnaire based on the WHO Responsiveness model. *PLoS One* 2015;10:e0117031.
71. Scheerhagen M, van Stel HF, Tholhuijsen DJ, Birnie E, Franx A, Bonsel GJ. Applicability of the ReproQ client experiences questionnaire for quality improvement in maternity care. *PeerJ* 2016;4:e2092.
72. Ayoubi S, Pazandeh F, Simbar M, Moridi M, Zare E, Potrata B. A questionnaire to assess women's perception of respectful maternity care (WP-RMC): development and psychometric properties. *Midwifery* 2020;80:102573.
73. Uludağ E, Mete S. Development and testing of women's perception for the scale of supportive care given during labor. *Pain Manag Nurs* 2015;16:751–8.
74. Bryant MJ, Schubert JP, Black RJ, Hill CL. Patient-Reported Experience Measures in outpatient rheumatology care: a systematic review. *Rheumatol Adv Pract* 2021;5:rkab079.
75. Cornelis C, den Hartog SJ, Bastemeijer CM, Roozenbeek B, Nederkoorn PJ, Van den Berg-Vos RM. Patient-reported experience measures in stroke care: a Systematic Review. *Stroke* 2021;52:2432–5.
76. Organization for Economic Co-operation and Development. Measuring what matters: the patient-reported indicators survey. Patient-Reported Indicators for Assessing Health System Performance. 2019. Available at: <https://www.oecd.org/health/health-systems/Measuring-what-matters-the-Patient-Reported-Indicator-Surveys.pdf>. Accessed October 10, 2022.
77. Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ Psychol Meas* 1973;33:613–9.
78. Male L, Noble A, Atkinson J, Marson T. Measuring patient experience: a systematic review to evaluate psychometric properties of patient reported experience measures (PREMs) for emergency care service provision. *Int J Qual Health Care* 2017;29:314–26.
79. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. *Control Clin Trials* 1991;12:142S–58S.
80. Kang S. Assessing responsiveness of the EQ-5D-3L, the Oxford Hip Score, and the Oxford Knee Score in the NHS patient-reported outcome measures. *J Orthop Surg Res* 2021;16:18.
81. Zygar-Hoffmann C, Schönbrodt FD. Recalling experiences: looking at momentary, retrospective and global assessments of relationship satisfaction. *Collabra Psychol* 2020;6:7.
82. Hastie R, Dawes RM. Rational choice in an uncertain world: the psychology of judgment and decision making. 2nd ed. SAGE Publications, Inc.; 2010
83. Huang WY, Wong SH. Cross-cultural validation. In: Michalos AC, ed. *Encyclopedia of Quality of life and well-being research*, Netherlands: Springer; 2014:1369–71.
84. Putnick DL, Bornstein MH. Measurement invariance conventions and reporting: the state of the art and future directions for psychological research. *Dev Rev* 2016;41:71–90.