medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Measurement of changes to the menstrual cycle: A transdisciplinary systematic review evaluating measure quality and utility for clinical trials

Amelia C.L. Mackenzie[1*], Stephanie Chung[2,3], Emily Hoppes[2], Alexandria Mickler[4#a], Alice Cartwright[2,3#b]

[1] Global Health and Population, FHI 360, Washington, District of Columbia, United States

[2] Global Health and Population, FHI 360, Durham, North Carolina, United States

[3] Department of Maternal and Child Health, University of North Carolina Gillings School of Global Public Health, Chapel Hill, North Carolina, United States

[4] Research, Technology, and Utilization Division, Office of Population and Reproductive Health, Bureau for Global Health, United States Agency for International Development and the Public Health Institute, Washington, District of Columbia, United States

[#a] Current address: Independent Consultant, Washington, District of Columbia, United States

[#b] Current address: Guttmacher Institute, New York, New York, United States

* Corresponding author:

Amelia Caroline Louise Mackenzie, PhD ScM

Scientist, Global Health and Population

FHI 360

2101 L St NW, Suite 700

Washington, DC 20037 USA

+1 919-321-3676

amackenzie@fhi360.org

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

ABSTRACT

Despite the importance of menstruation and the menstrual cycle to health, human rights, and sociocultural and economic wellbeing, the study of menstrual health suffers from a lack of funding, and research remains fractured across many disciplines. We sought to systematically review approaches to measure four aspects of changes to the menstrual cycle—bleeding, blood, pain, and perceptions—caused by any source and used in any field. We searched MEDLINE, Embase, and four instrument databases and included peer-reviewed articles published between 2006 and 2023 that reported on the development or validation of instruments assessing menstrual changes using quantitative or mixed-methods methodology. We evaluated instruments on measure quality and utility for clinical trials. From a total of 8,490 articles, 8,316 were excluded (i.e., 376 duplicates, 7,704 during title/abstract screening, and 236 during full text review), yielding 94 instruments from 174 included articles. Almost half of articles were from the United States or United Kingdom and over half of instruments were only in English, Spanish, French, or Portuguese. Most instruments measured bleeding, pain, or perceptions, but few assessed blood. Nearly 60% of instruments were developed for populations with menstrual or gynecologic disorders or symptoms. Most instruments had fair or good measure quality and/or clinical trial utility; however, most instruments lacked evidence on responsiveness, question sensitivity and/or transferability, and only three instruments had good scores of both quality and utility. Although we took a novel, broad, and transdisciplinary approach, our systematic review found important gaps in the literature and instrument landscape and a need to examine the menstrual cycle in a more comprehensive, inclusive, and standardized way. Our findings can inform the development of new or modified instruments, which—if used across the many fields that study menstrual health—can contribute to a more systemic and holistic understanding of menstruation and the menstrual cycle.

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Introduction

Menstrual health across disciplines

Menstruation and the wider menstrual cycle play a notable role in the health, human rights, and sociocultural and economic wellbeing of people who menstruate [1]. In addition, although its significance should not be utilitarianly reduced to only reproductive function, continuity of the human species would not occur without the menstrual cycle. Despite its importance, the study of menstruation and the menstrual cycle continues to suffer from a historical lack of funding and research across disciplines, including within the biological, clinical, public health, and social sciences. Within biomedical research, for example, a publication reporting on a recent technical meeting on menstruation convened by the United States (US) National Institutes of Health (NIH) decried a "lack of understanding of basic uterine and menstrual physiology" among researchers [2]. Indeed, many foundational, field-defining works have only recently emerged in the past five to ten years following increased attention to menstrual health, which the Global Menstrual Collective defined in 2021 as "a state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity, in relation to the menstrual cycle" [3]. The contemporary growth of the menstrual health field is—at least partly—due to grassroots menstrual activism, which resulted in 2015 being labeled as "the year of the period" in the lay press [4]. Other examples of recent fundamental work within menstrual health across disciplines include recommendations for the menstrual cycle to be considered a vital sign and the advent of the field of critical menstruation studies [5,6]. Despite these recent efforts, insufficient research on menstrual health persists. In addition, the study of menstrual health remains fractured across many fields and disciplines, many of which are siloed despite adjacent or even overlapping subject matters (e.g., menstrual health and hygiene within wider sexual and reproductive health; or gynecology, endocrinology, and many other specialties within medicine) [7,8]. As a result, we still lack a complete, systemic, and holistic understanding of menstruation and the wider menstrual cycle.

The type of interdisciplinary, comprehensive global efforts needed to address such large gaps in menstrual health research can greatly benefit from standardization—of terminology, of measurement, of analysis, and of outcomes or indicators. The widest global effort at standardization to date has taken place within medicine; the International Federation of Gynecology and Obstetrics (FIGO) established clinical standards of normal and abnormal uterine bleeding (AUB) occurring outside of pregnancy via a consensus-building process over a series of years [9–12]. These FIGO standards dictate four parameters

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

for menstrual bleeding: the frequency, duration, volume, and regularity of bleeding, with *normal* defined as bleeding occurring every 24-38 days (frequency), bleeding lasting no more than 8 days (duration), bleeding of a 'normal' amount as defined by the patient that does not interfere "with physical, social, emotional, and/or material quality of life" (volume), and bleeding within a menstrual cycle that only varies in length by plus or minus 4 days (regularity). FIGO defines bleeding outside these defined norms as *abnormal*, with AUB divided into standard categories based on whether it is acute or chronic and the source or etiology of the abnormality according to the acronym PALM-COEIN (i.e., Polyp, Adenomyosis, Leiomyoma, Malignancy and hyperplasia, Coagulopathy, Ovulatory dysfunction, Endometrial disorders, Iatrogenic, and Not otherwise classified). Other examples of efforts at standardization include menstrual hygiene indicators within the Water, Sanitation and Hygiene (WASH) field and defining how contraception can impact the menstrual cycle and analyzing these data in contraceptive studies [13–18].

Related to terminology, this review uses the phrase, "people who menstruate", which we define as those who can menstruate, do menstruate, or have menstruated. Although people who menstruate may or may not identify as women or girls, and not all women and girls menstruate [19], we do use the terms 'women' and 'girls' in some instances, especially when citing primary literature and because menstrual health cannot "be adequately addressed without attention to the gender norms and dynamics experienced by individuals in the cultures and communities in which they live" [7]. As much as possible, however, we use gender inclusive terms and other people-first language.

Review aim and scope

To aid in efforts for standardized measurement across menstrual health research, we sought to systematically review approaches that have been developed and validated to measure four aspects of changes to the menstrual cycle: bleeding, blood, pain, and perceptions of bleeding, blood, or pain. We use the term '**menstrual changes**' to refer to these four aspects for the remainder of the paper. Related reviews have been conducted: (a) within fields such as menstrual hygiene or the study of heavy menstrual bleeding (HMB) [20,21]; (b) to measure single parameters like volume of menstrual blood loss [22]; and (c) for specific approaches like pictorial methods to diagnose HMB [23]. However, given the gaps and silos within menstrual health research, our aim was to conduct an expansive and transdisciplinary review to inform standardized measurement across the study of menstruation and the menstrual cycle.

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

This broad approach resulted in two decisions about our review scope. First, we sought to include menstrual changes caused by any etiology or source. We are not aware of any previous efforts to look at menstrual changes across disciplines in this way, but there are many factors that can result in menstrual changes, including those endogenous and exogenous to the person who menstruates. Examples of these etiologies or sources include menstrual or gynecologic disorders like adenomyosis, use of hormonal or intrauterine contraceptives, use of other drugs or devices to treat or prevent disease, environmental exposures, infectious disease, injury, coagulation disorders, and diet and exercise. Our second decision on scope was to include any measures or methods for assessing menstrual changes. Examples of these measures or methods could include quantitative and semi-quantitative assays, biomarkers, or data reported by clinicians, researchers or directly by the person who menstruates. We use the term '**instruments**' to refer to any of these measures or methods for the remainder of the paper.

Clinical trial context

Although our broad approach does not preclude the use of our results to inform the measurement of menstrual changes across research contexts, one area for which we intend our review to be quite relevant is for data collection in clinical trials. Our immediate use of the review results is for the purpose of improving and standardizing the measurement of menstrual changes in clinical trials, specifically contraceptive clinical trials. The importance of data on menstrual changes in the clinical trial context was recently highlighted during the introduction of COVID vaccinations. Because vaccine trials did not collect data on any impact to the menstrual cycle or menopausal uterine bleeding, there were concerns among vaccinated people who menstruate when they experienced these changes, which can erode trust in clinical research and public health interventions [24–28].

Clinical trials, and the preclinical research that precedes them, collect data on key organ functioning and vital signs as part of standard toxicology and pharmacodynamics, yet data on the menstrual cycle are not routinely collected. Trials typically reflect the people, priorities, and purposes of those within the clinical trial ecosystem—that is, the individuals and systems that fund clinical research, conduct clinical trials, and regulate the drugs and devices tested in trials, as well as the individuals who participate in trials. Historically, there has been an underrepresentation of people who menstruate within the clinical trial ecosystem [29]. This exclusion is true for much of the preclinical research across many biomedical fields as well, and even cell lines used inting *in vitro* studies are predominantly derived from male animals [30,31]. Although proof-of-concept studies for drugs or devices intended for use in women, such

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

as contraceptives that may impact the menstrual cycle, do typically use female animals when the model organism used has an estrous or menstrual cycle, other preclinical research disproportionately rely on only male animals. Using both female and male animals, however, could provide early indications of any impacts on cycles, as well as many other sex-specific effects or differences. Despite decades of concrete efforts, sex and gender disparities persist in the clinical trial ecosystem [32–34].

Another element to the current clinical trial context relevant to our review is the increasing use of patient-reported outcomes (PROs). NIH and the US Food and Drug Administration (FDA) defines PROs as "a measurement based on a report that comes directly from the patient (i.e., study subject) about the status of a patient's health condition without amendment or interpretation of the patient's response by a clinician or anyone else," noting "a PRO can be measured by self-report or by interview provided that the interviewer records only the patient's response" [35]. PROs can include "symptoms or other unobservable concepts known only to the patient (e.g., pain severity or nausea) [that] can only be measured by PRO measures," as well as "the patient perspective on functioning or activities that may also be observable by others" [35]. Unless an assay or biomarker are used, all outcomes on menstrual changes are reported by the person who menstruates and, therefore, are PROs. The FDA has a series of guidance documents on the use of PROs and other clinical outcome assessments in clinical trials as part of patient-focused drug development efforts [36–39].

Review questions and objective

Given the aims of the review, our review questions were: (a) What instruments have been developed to assess menstrual changes caused by any etiology or source? and (b) What is the quality of these instruments and their utility for clinical trials? The objective of our systematic review was to compile a complete list of instruments used to measure menstrual changes and assess their quality and clinical trial utility.

MATERIALS AND METHODS

We conducted our systematic review in alignment with Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines [40–42]. We include a completed PRISMA checklist for this review in Table S1.

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Systematic review protocol

We developed a protocol per PRISMA guidance, and protocol drafts were reviewed by experts in the fields of menstruation and contraception who are members of the Global Contraceptive-Induced Menstrual Changes (CIMC) Task Force [14]. We registered our review protocol in PROSPERO (ID: CRD42023420358) [43].

Search strategy

We conducted a multi-stage literature search in collaboration with the FHI 360 health sciences library to identify peer reviewed articles examining instruments to measure menstrual changes. First, we conducted preliminary searches in MEDLINE to refine our search strategy, including PubMed search terms recommended by the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) [44]. We then reviewed the 50 most relevant hits from the Embase, CINAHL, and PsycINFO databases to determine which should be included in our search strategy in addition to MEDLINE. Only Embase contained relevant articles within those 50 most relevant hits, so it was the only other database included in our final search. Table 1 shows the final search strategy for MEDLINE, which included largely Medical Subject Headings (MeSH) Major Topic terms and title or abstract search terms. The MEDLINE search strategy was adapted by an FHI 360 health sciences librarian for Embase (Table S2). Final searches of MEDLINE and Embase were conducted, and the resulting records were uploaded into Covidence [45].

Table 1. MEDLINE search strategy

("menstrual cycle"[MeSH Major Topic] OR "menstruation disturbances"[MeSH Major Topic] OR "Endometriosis"[MeSH Major Topic] OR "Uterine Diseases"[MeSH Major Topic] OR "menstrua*"[Title/Abstract] OR "menses"[Title/Abstract] OR "uterine bleeding"[Title/Abstract] OR "vaginal bleeding"[Title/Abstract] OR "amenorrhea"[Title/Abstract] OR "dysmenorrhea"[Title/Abstract] OR "menorrhagia"[Title/Abstract] OR "oligomenorrhea"[Title/Abstract] OR "metrorrhagia"[Title/Abstract] OR "hypermenorrhea"[Title/Abstract] OR "hypomenorrhea"[Title/Abstract] OR "polymenorrhea"[Title/Abstract])
AND
("Surveys and Questionnaires"[MeSH Major Topic] OR "Pain Measurement"[MeSH Major Topic] OR "Patient Reported Outcome Measures"[MeSH Major Topic] OR "psychometrics"[MeSH Major Topic] OR "Sensitivity and Specificity"[MeSH Major Topic] OR "Validation Study"[Publication Type] OR "Validation Studies as Topic"[MeSH Major Topic] OR "measur*"[Title] OR "method*"[Title] OR "questionnaire*"[Title] OR "scale"[Title] OR "tool*"[Title] OR "patient reported outcome measure*"[Title/Abstract] OR "psychometr*"[Title/Abstract])
AND

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

("2006/01/01"[Date - Publication] : "2023/10/05"[Date - Publication])

Next, we searched four instrument databases for any relevant instruments measuring menstrual changes: (a) the NIH Common Data Element (CDE) Repository [46], (b) the COSMIN database of systematic reviews of outcome measurement instruments [47], (c) the Core Outcome Measures in Effectiveness Trials (COMET) Database [48], and (c) ePROVIDE databases [49]. We detail search strategies for these instrument databases in Table S2. Articles for any relevant instruments identified via these databases were uploaded into Covidence. We also planned to include instruments identified from searches of ClinicalTrials.gov and the Patient-Reported Outcomes Measurement Information System (PROMIS) database of measures, but multiple search strategies did not yield results we could screen and include.

Following screening and review of articles from the two literature databases (i.e., MEDLINE and Embase) and the four instrument databases (i.e., NIH CDE, COSMIN, COMET, and ePROVIDE), we completed two additional steps: (a) we extracted primary articles published since 1980 from all relevant review articles identified from the literature and instrument databases; and (b) we identified any original development articles for instruments developed before 2006. These primary articles and original development articles were then uploaded into Covidence for screening. Book chapters were excluded at this stage of screening.

Overall, our goal was to include all articles published on the (a) development, (b) validation, or (c) review of instruments since January 1, 2006. For instrument development or validation (a and b), we selected 2006 because the last major revision of standardized CIMC measurement in contraceptive clinical trials was published in 2007; therefore, that revision would encompass instruments developed or validated prior to 2006. For instruments reviewed (c), we selected 1980 as our date limit for extracting primary papers from identified reviews because the initial efforts to standardize CIMC measurement in contraceptive clinical trials, led by the World Health Organization (WHO), were in the 1980s; therefore, that WHO work would already encompass literature before 1980.

After completing our systematic review, we conducted an updated search to ensure the results reported up-to-date findings. Our original literature database search covered January 2006 through June 2022, and the updated search covered June 2022 through October 2023. For all identified articles in both searches, we completed the same search processes described above and the same screening and review

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

processes described below. The paper reports on total results from both searches combined, but we provide additional details and PRISMA diagrams for the individual searches in Supplementary file S3.

Inclusion/exclusion criteria

We included all peer-reviewed articles—including those with prospective, retrospective, or cross-sectional study designs, and review papers—that met our inclusion and did not meet our exclusion criteria. We detail these criteria in Table 2, but briefly, we included articles that: (a) reported on the development or validation of instruments to measure menstrual changes, (b) used mixed methods or quantitative approaches, and (c) were published between January 1, 2006 and October 5, 2023. We did not impose any restrictions on article language, country, or geographic region. Articles using only qualitative methods and conference abstracts, editorials, and commentaries were excluded because they would not contain the information necessary to evaluate instrument quality and utility for clinical trials, per our second review question.

Table 2: Inclusion and exclusion criteria

Inclusion criteria	1. Articles primarily focused on developing, validating, and/or evaluating instruments measuring menstrual changes or perceptions of menstrual changes, with information reported to assess instrument and/or study quality 2. Articles published between January 1, 2006 and October 5, 2023 3. Articles published in any language 4. Articles from any geographic region
Exclusion criteria	1. Articles with only qualitative data 2. Articles that were conference abstracts, editorials, and commentaries 3. Articles whose primary purpose was noy validating instruments measuring menstrual change, such as studies focusing on biomarkers or biological pathways of menstrual changes, cancer screening instruments, or studies of social-behavioral correlates of menstrual changes 4. Articles reporting only on data from people in menopause

Our definition of menstrual changes was adapted and broadened from the Global CIMC Task Force definition of changes to the menstrual cycle caused by contraception [14]. For the purposes of this review, the term, **menstrual changes**, includes four aspects (a) bleeding duration, volume, frequency, and/or regularity/predictability; (b) blood consistency, color, and/or smell; (c) pain or cramping; and (d) perceptions of bleeding, blood, or pain. We define perceptions as the perspectives on, attitudes about, experiences with, and acceptability of menstrual changes at the individual-level, interpersonal-level, community-level, and wider levels, including social norms. Examples of these four aspects of menstrual changes are: (a) an increase in how long bleeding lasts (bleeding duration), (b) a reduction of clotting

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

(blood consistency), (c) a decrease in dysmenorrhea (pain), and (d) an impact on quality of life or attitudes (perceptions of changes).

We use the single term '**instrument**' to capture any measure, method, or approach to assess menstrual changes, including healthcare provider-reported, menstruator-reported, researcher-based, biomarker-based, or assay-based methods, and including those that may be deemed "objective" or "subjective" and both directly observable and personal perceptions of menstrual changes (adapted from [50]). Our definition of **development** or **validation** of instruments was intentionally broad, including any manner of validation or evaluation (e.g., reporting any evidence on validity, reliability, responsiveness, interpretability, and other attributes of measure quality or utility) and any development or validation informed by input from research participants who menstruate.

Developing data extraction forms and instrument evaluation

One author (SC) drafted the initial template data extraction form in Excel after input from the rest of the authors, and all authors reviewed and gave feedback on the draft data extraction form. The final data extraction form collected information in five areas: article information, study design and sample information, details on the instrument, measure quality attributes, and clinical trial utility attributes. Table S4 has details on the fields of the data extraction form for each of the five areas.

For assessing measure quality and clinical trial utility, one author (SC) reviewed existing evaluation criteria and tools from the literature and guidance documents on selecting instruments for clinical trials (e.g., see Crossnohere *et al*., 2021 [51] for a recent overview) with input from the rest of the authors. After considering several alternatives (e.g., COSMIN Risk of Bias checklist [52], Francis *et al*.'s checklist to operationalize measurement characteristics of PRO measures [53], and the International Professional Society for Health Economics and Outcomes Research (ISPOR) PRO Good Research Practices Task Force guidance [54,55]), we determined these approaches did not meet our needs due to being too burdensome, too binary, or not specific to evaluation, respectively. We decided to follow the Patient-Reported Outcomes Tools: Engaging Users and Stakeholders (PROTEUS) Consortium recommendations to use International Society for Quality of Life Research (ISOQOL) standards for PRO measures [56,57]. We made two adjustments to the ISOQOL standards: (a) we added an attribute on sensitivity of questions given the topic of menstruation has a noted amount of stigma surrounding it [58]; and (b) we separated out participant burden from investigator burden given these two can differ greatly for instruments measuring menstrual changes. We categorized six attributes as related primarily to the

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

quality of the instrument (i.e., **measure quality**: conceptual/measurement model, reliability, content validity, construct validity, responsiveness, and sensitive nature of questions) and four attributes as related primarily to the utility of the instrument in clinical trials (i.e., **clinical trial utility**: interpretability of results, the transferability of the instrument, participant burden, and investigator burden).

We scored each attribute of measure quality and clinical trial utility on a scale from 0 to 3, where 0 indicated there were **no data** reported on the attribute, 1 indicated **poor** measure quality/clinical trial utility of the attribute, 2 indicated **fair** measure quality/clinical trial utility of the attribute, and 3 indicated **good** measure quality/clinical trial utility of the attribute. Criteria for scoring of an attribute was defined in line with ISOQOL standards [57] and reviewed by measurement and clinical experts at FHI 360 and within the Global CIMC Task Force. We show the measure quality and clinical trial utility attributes and scoring criteria in Table 3.

Table 3: Measure quality and clinical trial utility scoring criteria*

Attribute	Poor quality (1)	Fair quality (2)	Good quality (3)
Measure quality			
Conceptual and Measurement Model Definition: The conceptual model provides a description and framework for targeted construct(s) in the measure. The measurement model maps individual measure items to the construct(s). *Score 0 if not assessed in article.*	Minimal discussion of conceptual model or measurement model that maps measure items to the construct(s). Or minimal discussion of intended population or context for measure use.	Some discussion of conceptual and/or measurement model that maps measure items to the construct(s). Or some discussion of intended population and/or context for measure use.	Clearly defines and describes concept(s) included in model and intended population(s) and context for measure use. Or clearly describes how concept(s) are organized into measurement model, including evidence for dimensionality of the measure, how items relate to each measured concept, and the relationship among concepts.
Reliability Definition: The degree to which a measure is free from measurement error. *Score 0 if not assessed in article.*	There is minimal evidence for measure reliability (e.g., internal consistency reliability, test-retest reliability, or item response theory)	Unclear or unjustified methodology used for assessing reliability. Or, if used, reliability Cronbach α <0.70 for group-level comparisons without justification.	Methodology for collecting data is justified (e.g., a multi-item measure is assessed for internal consistency reliability and a single-item measure is assessed by test-retest reliability or item response theory).

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Attribute	Poor quality (1)	Fair quality (2)	Good quality (3)
			Or, if used, reliability Cronbach α ≥0.70 for group-level comparisons. If lower, there is clear and appropriate justification.
Content Validity Definition: The extent to which the measure includes the most relevant and important aspects of a concept in the context of a given measurement application. *Score 0 if not assessed in article.*	Minimal evidence participants or experts consider the measure relevant and comprehensive. Or minimal documentation of methodology for evaluating content validity.	Some evidence participants and experts consider the measure relevant and/or comprehensive for the concept, population, and/or intended application. Or some evidence of methodology used to evaluate content validity. Or the paper mentions past validation research (i.e., focus groups, pilot studies, formative research) but does not provide detail on these studies.	Clear evidence participants and experts consider the measure relevant and comprehensive for the concept, population, and intended application. And clear evidence of methodology used to evaluate content validity, including for assessing the relevance of measured concept(s), comparing validation study sample to the wider target population, and justification for recall period.
Construct Validity Definition: The degree to which scores on the measure relate to other measures (e.g., patient-reported or clinical indicators) in a manner that is consistent with theoretically derived a priori hypotheses concerning the concepts being measured. *Score 0 if not assessed in article.*	Minimal evidence supporting pre-determined hypotheses related to construct validity.	Some evidence supporting pre-determined hypotheses related to construct validity.	Clear evidence supporting pre-defined hypotheses on the expected associations among other measures similar or dissimilar to the studied measure.
Responsiveness/dynamism Definition: The extent to which a measure can detect changes in the construct being measured over time. *Score 0 if not assessed in article.*	Minimal evidence the measure can detect changes consistent with pre-defined hypotheses related to responsiveness. Or minimal evidence the measure can detect changes within or among participant groups.	Some evidence the measure can detect changes consistent with pre-defined hypotheses related to responsiveness. Or some evidence the measure can detect changes within or among participant groups.	Clear evidence the measure can detect changes consistent with pre-defined hypotheses in the target population for the intended application. And clear evidence the measure can detect changes within or

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Attribute	Poor quality (1)	Fair quality (2)	Good quality (3)
			among participant groups.
Sensitive nature of items Definition: How measure addresses questions of sensitive topics, including those that are seen as intrusive, posing a threat of disclosure, or eliciting socially desirable answers. *Score 0 if not assessed in article.*	Minimal evidence about measure or item sensitivity Or evidence of sensitivity that may result in biased responses	Some evidence or discussion about measure or item sensitivity Or some evidence of reduced sensitivity that would not result in biased responses	Clear evidence about measure or item sensitivity And clear evidence of reduced sensitivity that would not result in biased responses
Clinical trial utility			
Interpretability of results Definition: The degree to which one can easily understand a measure's results (e.g., scores, levels). *Score 0 if not provided in article.*	Minimal evidence for interpreting results. Or minimal evidence results are understood by relevant stakeholders. There is no clinically relevant minimum change or no assessment of clinical relevance.	Some evidence for interpreting results. Or some evidence results are understood by relevant stakeholders, including patients, clinicians, and/or researchers. There is an agreement on clinically relevant minimum change and/or assessment of clinical relevance.	Clear evidence of interpreting results, including differentiating between differing outcomes (e.g., high and low scores), and/or what constitutes a large or small change in the measured concept. And evidence results are clearly understood by multiple relevant stakeholders, including patients, clinicians, and researchers. There is an accepted clinically relevant minimum change.
Transferability Definition: The degree to which the measure can be transferred between linguistic and sociocultural groups. *Score 0 if not provided in article.*	Minimal evidence measurement properties are maintained across linguistic and/or cultural groups.	Some evidence measurement properties are maintained across linguistic and/or cultural groups.	Clear evidence measurement properties are maintained across linguistic or cultural groups, including qualitative testing of the translated measure.
Participant Burden Definition: The time, effort, resource (e.g., use or ownership of smart phone, internet access refrigeration), and other demands placed on those to whom the measure is administered. *Score 0 if not provided in article.*	Measure requires more than 20 minutes† to complete (>40 questions), requires data collection daily or multiple times a day, and/or multiple clinic visits or daily data collection outside the home. Or there is no	Measure requires between 15-20 minutes† to complete (20-40 questions), and/or one or two clinic visits, including those that are a burden to participant. Or there is limited information on expected participant	Measure requires less than 15 minutes† to complete (<20 questions), no daily data collection, and no more than one clinic visit. Or there is an accurate description of the expected participant time

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Attribute	Poor quality (1)	Fair quality (2)	Good quality (3)
	information on expected participant time burden. Or the measure requires resources not available to most participants. Or there is minimal information on literacy demand of measure items or appropriateness for proposed context.	time burden, including limited or no input from participant review panels. Or the measure may require some resources can be a barrier to some participants. Or literacy demand of measure items is above a 6th grade level (i.e., >12-year-old) and not appropriately justified for proposed context.	burden with approval from participant review panels. Or there are no resource barriers to participants. And literacy demand of measure items is at a 6th grade level or lower (i.e., ≤12-year-old), or literacy level is appropriately justified for proposed context.
Investigator Burden Definition: The time, effort, resource, and other demands placed on those who administer the measure. *Score 0 if not provided in article***.**	There is a high burden on the data collection team due to: (a) data collector training being time or cost prohibitive with a lack of available training materials; (b) a high data monitoring burden to maintain quality data; (c) measure scoring being complex; or (d) measure inflexible or resource intensiveness (e.g., can only be interviewer-administered or requires tablet or computer). Or there is minimal information on investigator burden.	There is a modest burden on the data collection team due to: (a) the time and cost of data collector training or lack of training materials; (b) data monitoring burden; (c) modest measure scoring complexity; or (d) the measure being either flexible or not resource intensive. Or there is limited information on investigator burden.	There is a low burden on a data collection team due to (a) minimal requirement for data collector training and availability of training materials; (b) low data monitoring burden, (c) measure scoring being simple, or (d) the measure being flexible and not resource intensive (e.g., either measure is completed by the participant or is easily explained and completed). Or there is an accurate description of the expected investigator burden.

* Attributes and definitions from Reeve *et al*. 2013 [57] per PROTEUS-Trials Consortium guidance [56], with modified as specified in the text.
† Crossnohere *et al.*, 2021 [51].

Process for title/abstract screening, full text review, and data extraction

The authors met with the FHI 360 health sciences library team for a month to finalize the search strategy and then began weekly author meetings to discuss progress, questions, and discordance, and to document decisions and progress in a shared Word document. We began title/abstract screening with an 'inter-reviewer reliability' meeting where all authors completed title/abstract screening on the same 50 articles to establish and confirm group standards. Then, two authors independently screened each

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

remaining title/abstract and two authors independently reviewed each relevant full text in Covidence. We resolved any discordance during weekly meetings via consensus conversations. We used the text translation feature of Google Translate to review abstracts not in English during screening, and we used the document translation feature of Google Translate and/or consulted a fluent colleague to review full text articles that were not in English. We used the notes and tag features in Covidence to document questions between meetings, consensus decisions during meetings, and any translation from Google Translate. We used Excel worksheets for data extraction. For instruments reported in more than one article, we concurrently extracted all articles on each instrument. We conducted data extraction with a fluent colleague for full text articles not in English. During title/abstract screening, full text review, and data extraction, when the authors had finished with approximately 5% of the articles, the following weekly author meeting included a specific discussion on the need for any clarifications or minor modifications to our inclusion/exclusion criteria for screenings/review or data extraction forms. After these '5% discussions', we made only minor clarifications to the inclusion/exclusion criteria and added or revised only a few fields in the data extraction forms.

Data analysis

Two authors (EH and SC) developed the initial analysis plan with input from the rest of the authors, and one author (EH) compiled all extracted data and conducted initial analyses with data checks by the rest of the authors. After data compilation, all authors conducted parts of the analysis. All analysis was conducted in Excel and included counts and frequencies, as well as specific analyses to assess (a) measure quality and (b) clinical trial utility. For these two outcomes, two authors (EH and SC) developed a scoring system with input from other authors in order to assign each instrument a measure quality score, a clinical trial utility score, and a total evidence score. For **measure quality scores** and **clinical trial utility scores**, we used an average of the highest score for each attribute of measure quality or clinical trial utility across all articles on an instrument. Because instruments could have more than one article providing data on measure quality and/or clinical trial utility and not every article evaluated all attributes, we did not include scores of zero (i.e., no data reported) in the measure quality and clinical trial utility scores. To reflect these differences in the number of articles and attributes reported in the article(s), we also calculated a total **evidence score**, which was the total of all scores—including zeros—across all attributes of measure quality and clinical trial utility. The total evidence scores, therefore, 'penalize' instruments for a lower level of evidence due to fewer articles or less attribute data and vice versa.

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

These three scores—measure quality (ranging from 1-3), clinical trial utility (ranging from 1-3), and total evidence (ranging 0+)—reflect different dimensions of an instrument. For example, two instruments might both have a score of 2.5 for measure quality, but one instrument might have an evidence score of 10 and the other, 100, indicating the latter has considerably more evidence and likely more certainty in the measure quality score. Alternately, two instruments may have similar measure quality and evidence scores, but one may have a clinical trial utility score of 1 and the other a score of 3, indicating the latter is likely better suited for use in clinical trials despite the similar levels of measure quality and evidence.

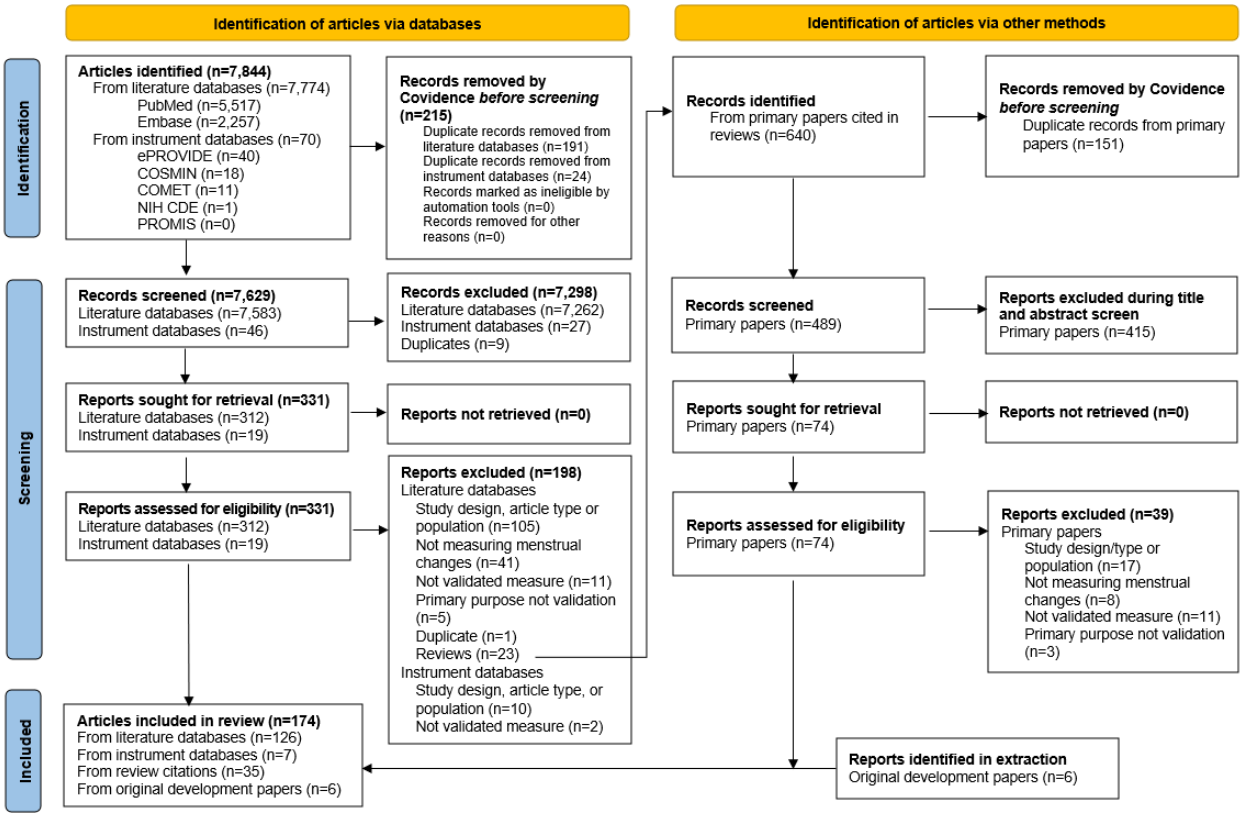
RESULTS

Search results

Our database searches yielded a total of 7,844 articles, of which 7,774 were from literature databases and 70 from instrument databases. Covidence removed 215 duplicates and we excluded 7,298 articles during title/abstract screening. During full text review, we excluded 115 articles for study design, article type, or population, 41 for not measuring menstrual changes, and 18 for no validation. We also identified one additional duplicate and found 23 relevant review articles. From these review articles, we extracted 640 primary articles, of which 35 remained after title/abstract screening and full text review. During data extraction, we identified 6 instruments for which we did not have the original development papers, because either they were developed before 2006 (i.e., our search strategy date limit; n=5) or had not been captured via our search strategy (n=1). Across all sources, our searches yielded 8,490 articles. We removed 376 duplicates, excluded 7,704 articles during title and abstract screening, and excluded 236 articles during full text review. In total, we identified 174 relevant full text articles of instruments developed, validated, or reviewed between January 1, 2006 and October 5, 2023. We present additional details on our search results and screening in the PRISMA diagram in Figure 1.

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Fig. 1. PRISMA Diagram*, **



* Per Page *et al*., 2021 [40]

**Across all searches depicted here, we identified a total of 8,490 articles (7,844+640+6). There was a total of 376 duplicates (191+24+9+1+151), and we excluded a total of 7,704 articles (7,262+27+415) during screening and a total of 236 articles (105+41+11+5+10+23+2+39) during full text review. We included a total of 174 (8,490-376-7,704-236) articles.

We found some similarities across papers that we excluded for not meeting our inclusion criteria. For example, we excluded conference presentations that never became full papers, studies that focused on validating instruments among only menopausal populations (e.g., [59,60], and studies that only validated surgical or treatment outcomes (e.g., [61,62]. In addition, there were two recent papers on core outcome sets for HMB and endometriosis relevant to the wider topic of measuring changes to the menstrual cycle, but we excluded them because there were no instrument details to extract [63,64].

Included article characteristics

Over 85% of the 174 articles were from either Europe (43%), North America (32%) or Asia (13%), and there were less than 15 articles from South America (n=13), from the Middle East (n=11), from Oceania

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

(n=8) and from Africa (n=5; note, some articles report data from more than one geographic region or more than one country, so the sum of article counts will be more than 174 and the sum of percents will be above 100). Just under half of articles were from only the United States (28%) or the United Kingdom (16%), although we did identify articles from a total of 50 countries. All articles were in English—even those reporting on instruments in other languages—except for two in Portuguese [65,66]. The most common study designs were cross-sectional or prospective cohort. We present details of all 174 included articles in Table S5.

Instrument characteristics

From the 174 included articles, we extracted 94 instruments. Almost three quarters (72%, n=68) were full instruments, collecting data on one or more menstrual change. Nearly a quarter (22.5%, n=21) were broader instruments that included sub-scales (8.5%, n=8) or a small number of items (14%, n=13) on menstrual changes. Five percent (n=5) were general instruments validated in menstruating populations on one or more menstrual change. The instruments with the most articles in our review were the Endometriosis Health Profile-30 (EHP-30; 20 articles), the Pictorial Blood Loss Assessment Charts & Menstrual Pictograms (PBAC; 11 articles), the Uterine Fibroid Symptom and Quality of Life questionnaire (UFS-QOL; 9 articles), the Polycystic Ovary Syndrome Quality of Life scale (PCOS-QOL; 8 articles), and the Endometriosis Health Profile-5 (EHP-5), Menstrual Attitudes Questionnaire (MAQ), and menstrual collection (5 articles each). In addition, about a third (38%, n=26) of full instruments used electronic data collection, and almost all full instruments (97%, n=66) were completed by only the patient/participant who menstruated (i.e., were PROs). We present the list of full instruments and instrument characteristics in Table 4, and details on the sub-scales, items, and general instruments are in Table S6.

Table 4: List of full instruments and characteristics

See end of file for Table 4.

Language(s)

Of the 68 full instruments, two-thirds were in English (66%, n=45), followed by Spanish (13%, n=9), French (9%, n=6), and Portuguese (9%, n=6); however, we identified instruments in 28 languages. About forty percent of instruments (41%, n=28) were only in English, although about a quarter of instruments (26%) were in more than one language, and six instruments were in at least 4 languages. These instruments included the EHP-30 (13 languages), UFS-QOL (5 languages), MAQ (5 languages), PBAC (4 languages), Endometriosis Daily Diary (EDD; 4 languages), and the Daily Diary (4 languages). We present

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

language details for all full instruments in Table 4 and for sub-scales, items, and general instruments in Table S6.

Specific Populations

Nearly 60% (n=40) of the 68 full instruments were developed and/or validated in populations with menstrual or gynecologic disorders or symptoms (i.e., 18 for endometriosis, 10 for HMB, 9 for dysmenorrhea, and 3 for uterine fibroids). Less than a quarter (24%, n=16) of full instruments were developed for and validated with adolescents (mean ages less than 18, n=10) or young people (mean ages early 20s, n=6). Three full instruments were specifically developed for those in perimenopause. A few instruments were developed or validated in populations of athletes or people in the military. No instruments or articles indicated inclusion of trans and gender nonbinary populations who menstruate.

Menstrual change(s) measured

Among the 68 full instruments, nearly half (46%, n=31) measured more than one of the four aspects of menstrual changes (i.e., bleeding, blood, pain, and perceptions). Nearly half of full instruments (49%, n=33) measured bleeding, about half (47%, n=32) measured uterine cramping or pain, and almost three quarters (74%, n=50) measured perceptions. Only eight (12%) measured blood. As shown in Table 4, three instruments assessed all of the four parameters of bleeding—duration, volume, frequency, and regularity/predictability (i.e., the Aberdeen Menorrhagia Severity Scale [AMSS], the New Zealand Survey of Adolescent Girls' Menstruation, and the World Endometriosis Research Foundation Endometriosis Phenome and Biobanking Harmonisation Project Standard Questionnaire [WERF EPHect EPQ-S]). No instrument assessed each of the three parameters of blood—color, consistency, and smell.

In addition, no instrument measured all parameters of menstrual changes, and only seven instruments measured at least a single parameter of each of the four aspects of menstrual changes. These instruments were the AMSS; electronic Personal Assessment Questionnaire - Menstrual, Pain, and Hormonal (ePAQ-MPH); Endometriosis Self-Assessment Tool (ESAT); Fibroid Symptom Diary (FSD); Menstrual Bleeding Questionnaire (MBQ); Menstrual Insecurity Tool; and the New Zealand Survey of Adolescent Girls' Menstruation. We present data on the menstrual changes measured for sub-scales, items, and general instruments in Table S6.

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

How instruments measured bleeding

Among the 33 full instruments measuring bleeding, most measured bleeding volume (n=26) and/or duration (n=12), while 8 instruments measured bleeding frequency and 9 measured bleeding regularity/predictability (Table 4). We describe the measurement of each parameter within full instruments in detail below.

We present sub-scale titles and item wording on bleeding in Table S6. There were 7 instruments with sub-scales that collected data on bleeding, 8 instruments with one to five items on bleeding, and two general instruments with items that asked about bleeding. Most sub-scales and items were for bleeding volume or regularity/predictability, often using terms not clearly defined or elaborated (e.g., 'regular' and 'normal').

Bleeding volume

Of the 26 full instruments that measured bleeding volume, 5 were only for volume and no other parameter of bleeding or other aspects of menstrual changes. Of those, 3 were instruments that semi-quantitatively measured blood volume via used menstrual products, including alkaline hematin assays and menstrual collection or record and recall measures. One instrument relied on respondents to estimate bleeding volume through the Mansfield-Voda-Jorgensen Menstrual Bleeding Scale, and one was a statistical model for estimating blood loss that was developed based on previously collected hematological values, daily diaries, and patient age among participants with HMB [67].

The 21 other instruments that measured volume also assessed other menstrual changes, and most (n=18) were designed for use by people with menstrual or gynecologic disorders and symptoms. Fourteen instruments were questionnaires, 5 were diaries, 1 used pictorial references, and 1 was a visual analog scale (VAS) where volume was rated on a scale from 0 (no bleeding) to 100 (the heaviest possible bleeding ever experienced). Most instruments asked about perceived volume of blood loss, usually by asking respondents to describe their bleeding in some range of light, medium, or heavy and/or reporting on the number of menstrual products (pads and/or tampons) they used on the heaviest day of their period. Some instruments also asked how many days of heavy bleeding the respondent experienced during the last cycle and how many days required double protection with multiple products at the same time. A few asked whether respondents had bleeding heavy enough to stain clothing or required getting up in the middle of the night to change menstrual products. PBAC and other similar pictorial assessments had respondents estimate the amount of bleeding via pictorials of

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

used pads and/or tampons. Of note, terms like 'light', 'heavy' and/or 'spotting' were not always or consistently defined across instruments, and there was a wide range for the frame of reference for recall, with diaries asking every day, other instruments asking about the last month or last menses/bleeding episode, and others asking more generally about experiences people typically have during menses/bleeding episodes.

Bleeding duration

The 12 full instruments that assessed bleeding duration did so in a variety of ways. One instrument used prospective diaries to record the first and last days of menses/bleeding episodes just to measure duration [68]. Three instruments measured duration and another parameter of bleeding, either using diaries and/or annual interviews [69,70] or a question on days of bleeding for every menstrual period over four months [71]. Eight instruments—seven questionnaires and one diary—measured bleeding duration along with other menstrual changes (i.e., blood, pain, or perceptions), and most (n=6) were developed for people with menstrual or gynecologic disorders and symptoms (e.g., HMB, endometriosis, or fibroids). The questionnaires generally asked respondents to note how many days their menses/ bleeding episodes last on average, either in general or in the last three months. Three instruments specifically asked if respondents had bleeding for more than seven days per month [72–74]. Finally, the diary asked respondents to note if they had bleeding on specific days [75].

Bleeding frequency

Of the 8 full instruments that measured bleeding frequency, 3 collected data only on frequency and no other parameter of bleeding or other aspects of menstrual changes. These asked respondents a few retrospective questions: "'How long is your menstrual cycle, on average? In other words, how many days are there from the first day of one menstrual period to the first day of the next period?" [76,77] or to recall the first date of their last menstrual period [78]. Another used a retrospective questionnaire on usual, shortest, and longest menstrual cycle length in the past 12 months, and this was compared to a prospective diary for two menses/bleeding episodes [79].

Five additional instruments asked about frequency along with other menstrual changes: a diary and 4 questionnaires. The questionnaires asked respondents to state how many days there were, on average, between the start or first day of one menses/bleeding episode to the first day of the next menses/bleeding episode with a 3-month recall period in two of the three questionnaires [80,81] or whether their menstrual cycle was between 21 and 45 days [82].

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Bleeding regularity/predictability

Nine instruments—all questionnaires—measured bleeding regularity/predictability as well as other changes, and about half of these (n=5) were specifically developed for those with menstrual or gynecologic disorders and symptoms. Almost all had respondents report if their bleeding was regular or irregular in general or over the past three months, but regularity was not defined further. One instrument—the MBQ—asked respondents if both their bleeding start and end dates in the last month were completely, somewhat, or not at all predictable [73], and another—the ePAQ-MPH—contained a regularity domain, which asked about both regularity of timing and predictability [83].

How instruments measured blood

Of the eight full instruments measuring blood, seven measured blood consistency (e.g., clotting), one measured blood smell, and none assessed blood color (Table 4). Full instruments that collected information about blood consistency were the PBAC/pictorial assessments, five questionnaires, and one diary. The questionnaires and diary specifically asked about blood clots—either ever or during the past month—while one also asked about "thick bleeding" during menstrual periods [84]. The one instrument that collected information about smell—the Menstrual Insecurity Tool—asked about smell of the "menstrual cloth, napkin, or [respondent's] body" [85].

We present details on sub-scales and items that measured blood in Table S6. One instrument had a subscale that collected data on blood color, consistency, and smell (i.e., the Menstrual Cycle-Related Signs and Symptoms Questionnaire subscale Section 1), and one other instrument had an item that asked about blood consistency (i.e., the Stellenbosch Endometriosis Quality of Life Measure).

How instruments measured pain

Of the 32 full instruments measuring pain, five measured only pain and no other menstrual change (Table 4). Two of these were VAS or numeric rating scales (NRS), where pain experienced was rated on a scale from 0 (no pain) to 10 or 100 (worst or unbearable pain). One instrument used a rubber bulb, which participants squeezed and corresponding measurements were recorded in reference to pain experienced [86], in another instrument participants were given a diagram of the body and asked to paint the areas affected by pain during their current menstrual period [87], and the final instrument included a single, retrospective question asking respondents to classify their frequency of menstrual discomfort as "always," "often," "sometimes," or "never" [88].

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

The remaining 27 full instruments measured pain in addition to other menstrual changes, and about two-thirds (n=19) were developed for use with those with menstrual or gynecologic disorders and symptoms, including 12 specifically for endometriosis. Seventeen of the instruments were questionnaires and eight were diaries. Ten used NRS measures, 8 asked about the use of and/or dosage of pain medications, 12 asked about whether pain affected daily activities or quality of life, and 11 asked about pain and sexual activity/vaginal penetration. Four instruments had extensive sections on pain, covering multiple aspects. These included the ePAQ-MPH [83], the Endometriosis Pain and Bleeding Diary [89], the New Zealand Survey of Adolescent Girls' Menstruation [90], and WERF EPHect EPQ-S [81].

Instruments with subscales (n=4), instruments with one to five items (n=7), and general instruments (n=3) also asked about pain (Table S6).

How instruments measured perceptions

Over half (n=29) of the 50 full instruments measuring perceptions about the impact of menstruation on life were developed for those with menstrual or gynecologic disorders and symptoms (Table 4). Most (n=41) full instruments were questionnaires and 9 were diaries. About three-quarters (n=38) assessed how aspects of the menstrual cycle impacted people's daily activities, including work, social/leisure activities, walking or sitting. About a third (n=15) of the full instruments asked specifically about pain limiting activities, and 19 asked more generally about the impact of menstruation or disorders on activities. Some instruments asked about the impact of multiple symptoms on activities. Over a third (n=16) of the full instruments asked about impact or limits on sexual activity, including general impact (n=7), from pain (n=11), or from bleeding (n=3). Here too, some instruments asked about the impact of multiple symptoms on sexual activity. About a quarter (n=13) of full instruments asked about the impact of menstrual changes on sleep, 7 on the general impact and 6 that were specific to pain. About two-thirds (n=32) of full instruments asked about emotions, either changes during the menstrual cycle or the impact of symptoms—such as in bleeding or pain—on their emotions. A few (n=6) full instruments had items on menstrual hygiene management, most of which were in low- and middle-income country settings [85,91–94].

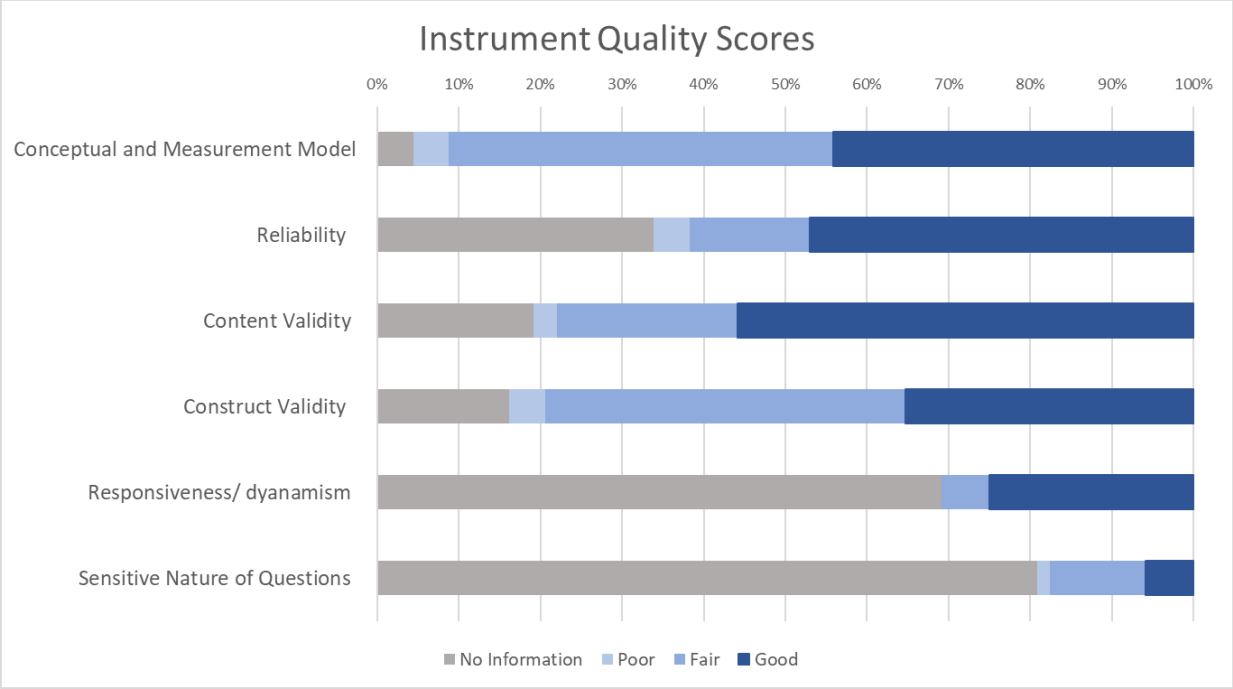
Instruments with subscales (n=2), instruments with one to five items (n=6), and general instruments (n=4) also asked about perceptions (Table S6).

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Measure quality of full instruments

When assessing measure quality (i.e., conceptual or measurement model, reliability, content validity, construct validity, responsiveness, and sensitive nature of questions), we found only five of the 68 full instruments (7%) had data on each of the six attributes of measure quality. These were the PBAC, EHP-30, Dysmenorrhea Daily Diary, MBQ, and a quantitative model for menstrual blood loss [67], each indicated by †† in Table 4. All but three instruments (96%, n=65) had evidence of a conceptual or measurement model and most also included evidence of content validity (81%, n=55), construct validity (84%, n=57) and reliability (66%, n=45); however, less than a third of instruments had evidence on responsiveness (31%, n=21), and less than a fifth had evidence on question sensitivity (19%, n=13, Figure 2).

Fig. 2. Instrument measure quality by attribute for full instruments



Of the 68 full instruments, 18% (n=12) had an overall **good** measure quality score (i.e., a score of 3), about three quarters (74%, n=50) had a **fair** measure quality score (i.e., a score less than 3 but greater than or equal to 2), and 9% (n=6) had a **poor** measure quality score (i.e., a score less than 2 but greater than or equal to 1; Figure 2). When we looked at individual attributes of measure quality, over half of instruments had a good score for content validity (56%, n=38), 47% had a good score for reliability (n=32), 44% had a good score for conceptual or measurement model (n=30), and over a third of instruments (35%, n=24) had a good score for construct validity; however, only a quarter had a good

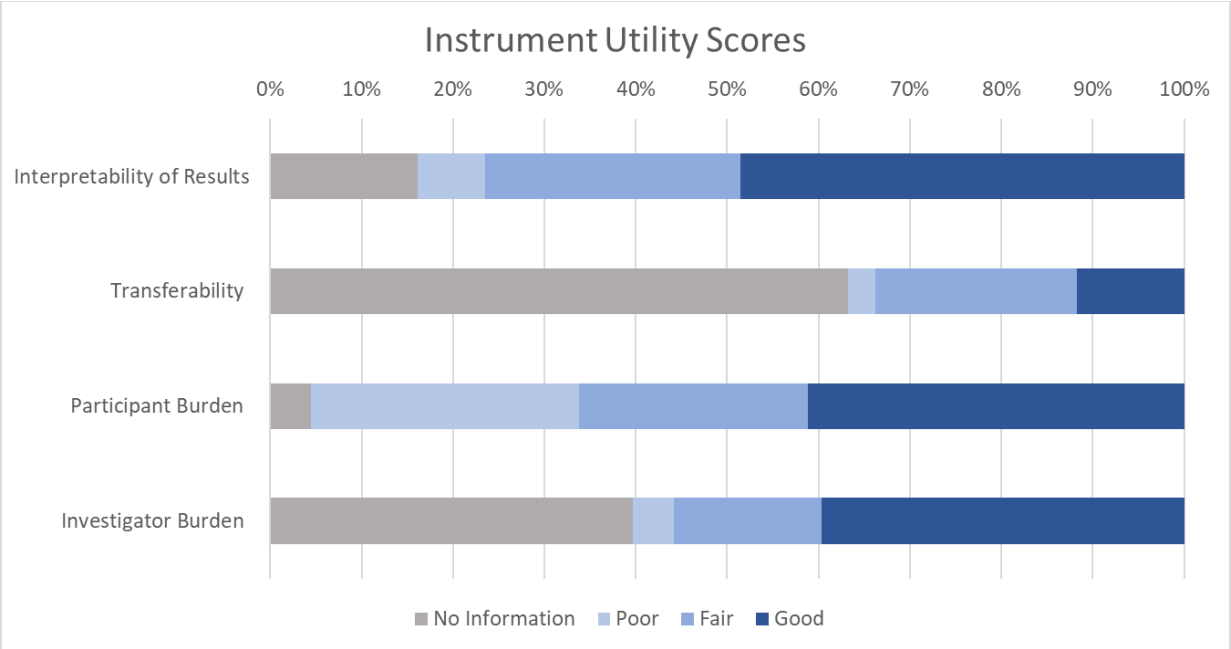
medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

score for responsiveness (25%, n=17), and only 4 instruments (6%) had a good score for question sensitivity.

Utility for clinical trials of full instruments

When assessing clinical trial utility (i.e., interpretability of results, transferability, participant burden, and investigator burden), we found 11 full instruments (16%) had data on each of the five attributes of utility, each indicated by ǂ in Table 4. All but three instruments (96%) had information on participant burden, 84% (n=57) had evidence of the interpretability of the instrument results, and slightly less than two thirds (60%, n=41) had documented investigator burden; however, only just over one third (37%, n=25) had evidence of transferability (Figure 3).

Fig. 3. Instrument utility in clinical trials by attribute for full instruments



Of the 68 full instruments, 22% (n=15) had an overall **good** clinical trial utility score, almost two thirds (62%, n=42) had a **fair** score, and 13% (n=9) had a **poor** score. When we looked at individual attributes of clinical trial utility, almost half of instruments (49%, n=33) had a good score for the interpretability of results, about 40% had good scores for participant burden (41%, n=28) or investigator burden (40%, n=27), but only 8 instruments (12%) had good scores for transferability (Figure 3).

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Overall full instrument evidence

Only the PBAC had evidence on all attributes of measure quality and all attributes of clinical trial utility, and only three instruments had both a good measure quality score and a good clinical trial utility score: EHP-5, the Spanish Society of Contraception Quality-of-Life (SEC-QOL), and the SAMANTA Questionnaire. Thirteen instruments had both measure quality scores and clinical trial utility scores greater than 2.5. Only one instrument, the Squeezing Pain Bulb, had both poor measure quality and poor clinical trial utility.

Full instrument total evidence scores ranged from 4 for the World Health Organization Disability Assessment Schedule 2.0 to 332 for the EHP-30, with an overall median score across instruments of 16 and mean score of 27 (Table 4). Overall, the following instruments had the five highest scores across measure quality, clinical trial utility, and total evidence: EHP-30, EHP-5, UFS-QOL, PBAC, and MBQ.

DISCUSSION

Our broad, interdisciplinary systematic review on the measurement of menstrual changes caused by any intrinsic or extrinsic factor, etiology, or source yielded 174 relevant articles and 94 instruments. Through our data extraction and analysis of these articles and instruments, we found several strengths and notable gaps in this literature around geographic and linguistic representation, how menstrual changes were measured, measure quality and clinical trial utility, and menstrual stigma, among others.

Geographic and linguistic representation

We identified articles from all geographic regions and 50 countries, and full instruments in 28 languages, including over a quarter in more than one language. Despite this evidence of the breadth of the literature, three quarters of articles were from North America or Europe and almost half were from just the United States and United Kingdom. In addition, over half of full instruments were only in English, Spanish, French, or Portuguese. These findings indicate the existing instrument landscape centers around the US and Western Europe, as well as colonial languages.

How menstrual changes were measured

We again found promising strengths mixed with important gaps when examining the menstrual changes that instruments measured and how they were measured. Although many full instruments measured perceptions, at least one parameter of bleeding, or pain, only 8 full instruments measured blood. It is

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

possible this lack of data collection on blood is due to the wide influence of menstrual stigma, especially the common perspective that menstrual blood is 'dirty' and requires 'hygiene' products to cleanse, absorb, and hide blood or odor [58,95,96]. No full instruments measured all parameters for each of the four aspects of menstrual changes we assessed, and only 7 instruments measured at least one parameter for all four. In addition, across all aspects of menstrual changes, there were not high levels of uniformity between instruments on how they measured each menstrual change, and many did not explain or define key terms (e.g., 'heavy', 'regular'), leaving their interpretation up to each respondent. This lack of clarity and specificity raises concerns about measurement error for a topic like menstruation and the wider menstrual cycle, around which there is high stigma and low health literacy and therefore, reduced shared understanding and references. These findings indicate there is a lack of instruments that examine all parameters and aspects of changes to the menstrual cycle in a comprehensive and standardized way.

Nearly 60% of full instruments we identified were developed for those with menstrual or gynecologic disorders and symptoms. In fact, the 3 instruments that accounted for almost a quarter of all identified articles—the EHP-30, PBAC, and USF-QOL—were each developed for use in populations with endometriosis, HMB, and fibroids, respectively. Instruments for these populations are of crucial importance, and it is encouraging to see over 70% of articles we identified published in the last 5 years study menstrual or gynecologic disorders and symptoms. However, the measurement of menstrual changes resulting from these disorders, such as very heavy bleeding and high levels of pain, may not translate to the menstrual changes experienced by the wider menstruating population or to the range of menstrual changes likely to occur across clinical trials and related research. For example, the extension of an instrument developed for those with HMB to a clinical trial of a hormonal contraceptive—which generally decreases bleeding volume—is yet to be supported by evidence. This difference is important because we could hypothesize, for example, there would be a difference in recall from a bleeding episode that resulted in stained clothing (i.e., from HMB) compared to a bleeding episode that did not interfere with daily activities (i.e., from a hormonal contraceptive). Because of these findings, instruments likely need to be developed or modified to capture a wider array of changes in bleeding, blood, and pain, as well as changes that are of smaller—but still meaningful—magnitude.

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Instrument quality and utility

From our assessments of measure quality and clinical trial utility for full instruments, we also found variability in our outcomes. Over 80% of instruments had either fair or good scores for measure quality and/or clinical trial utility, and only one had both poor measure quality and poor clinical trial utility. On the other hand, only three instruments had both good measure quality and good clinical trial utility.

We also note almost all instruments had evidence supporting some quality and utility attributes but not others. Sixty percent or more of instruments had evidence of a conceptual or measurement model, reliability, content validity, or construct validity for measure quality, or had evidence of interpretability of results, participant burden, or investigator burden for clinical trial utility; almost a quarter of instruments had evidence of each of these seven attributes. On the other hand, only one instrument—the PBAC—had evidence for all attributes of quality and utility, and over 60% of instruments did not have evidence of responsiveness, question sensitivity, or transferability, with nearly 40% not having evidence of any of the three. Each of these largely missing attributes are likely to be important for any instrument used broadly, especially in clinical trials. Such an instrument will need to: (a) capture changes during drug/device use (responsiveness); (b) not be viewed as too intrusive or stigmatizing (question sensitivity); and (c) be used in multiple linguistic and sociocultural contexts (transferability).

Menstrual stigma and other notable gaps

Our findings on the limited measurement for blood and lack of evidence for question sensitivity highlight the importance of menstrual stigma. We often found a contradiction during the development and validation of instruments; although menstrual stigma was frequently acknowledged as part of the sociocultural milieu surrounding menstruation, instruments generally did not adequately address menstrual stigma or how stigma may relate to question sensitivity and the potential impact of this on data quality or measurement error.

Beyond the difficulty of measurement due to menstrual stigma, there is innate complexity in measuring changes to a biological process that, itself, consists of so many facets that change over time and vary between individuals [97,98]. For example, there are changes within a single menstrual cycle (e.g., different bleeding and/or pain experienced on different days of a cycle), between menstrual cycles during the same year, and over the course of the menstruating life course of people who menstruate, as well as differences between individuals who menstruate [99–101]. These factors are important when we consider just under half of articles for the identified full instruments had cross-sectional study designs.

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

In fact, this study design limitation could be the reason we found a lack of evidence on instrument responsiveness and measurement of more temporally related parameters like bleeding frequency and regularity/predictability.

In addition to the gaps in the literature and instrument landscape already mentioned, three additional findings warrant attention. First, only just over a third of instruments used electronic data collection. Although this may be partly due to our review extending through 2006, given the data quality and monitoring benefits of electronic data collection and with the current proliferation of period tracking and other FemTech applications [102,103], new and refined instruments will likely need strong justification for not proceeding in this direction. In addition, there is a need to establish the equivalence between existing paper instruments and any electronic versions developed, ideally in accordance with established approaches like the ISPOR good research practices on use of mixed mode PROs [104].

Second, there is a lack of attention paid to the two ends of the menstruating life course. There were only ten instruments specifically developed with data from adolescents and three instruments developed for those in perimenopause, both groups who can experience an increased amount of variability and change in their menstrual cycles as compared to the middle of the menstruating years [105]. In addition, data on older menstruators were often collapsed for people who were in perimenopause and menopause/post-menopause, or age was commonly used as a proxy for this process and transition. Although the age range for menopause is narrower than that of menarche, given the general lack of research around menopause and the preceding and succeeding years, it seems the opposite should be true (i.e., more data and larger sample sizes among people around the end of their menstruating years is warranted) [106].

Third, we found a lack of inclusion for trans and gender nonbinary populations in all articles for all instruments. As we note in the introduction of this paper, people who menstruate may or may not identify as women or girls, and not all women and girls menstruate. It is important to engage all populations who menstruate in the development of instruments to measure changes to the menstrual cycle. Inclusion of sexual and gender minority (SGM) individuals who menstruate in clinical trials is a noted priority among NIH and other funders and researchers. In addition to NIH establishing its SGM Research Office in 2015, *clinical research* is the first theme of the current Strategic Plan to Advance Research on the Health and Well-being of SGM populations [107].

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Limitations of the review

Although we followed PRISMA guidelines and included 'inter-reviewer reliability' checks, weekly meetings, and multiple reviewers per article, there are a few limitations to note about our review process. The most important limitations are related to decisions made regarding the scope of the review to make it focused and feasible. First, we only included four aspects of changes to the menstrual cycle: changes in bleeding, blood, pain, and perceptions of bleeding, blood, or pain. Although these aspects are likely the most studied thus far, there are many other important changes to the menstrual cycle, including in hormone levels, the phases or characteristics of phases of the menstrual cycle, and other symptoms besides pain. As the study of menstrual health grows, it will be important for future reviews to consider these areas of research. Another limitation of our scope is the exclusion of other types of uterine bleeding outside of the menstrual cycle, such as bleeding during pregnancy, while breastfeeding, and after menopause. Future insights into how these types of bleeding relate to bleeding during the menstrual cycle will be important to our research and understanding of all uterine bleeding.

We also note a few limitations related to our review process. First, although all authors have training and experience across multiple disciplines, none are experts in all fields from which we drew our literature given our broad, interdisciplinary approach. We aimed to address this limitation by consulting other experts internally at FHI 360 and members of the Global CIMC Task Force when we encountered a question or issue outside of our knowledgebase, but it is still possible we missed articles, data for extraction, or other elements due to this limitation. Second, our primary use of the review was for the context of contraceptive clinical trials, so it is possible this internal aim may have biased our decisions about including or excluding articles. From the very beginning of the review, however, we had the aspiration for the review to be useful across contexts and disciplines, so our protocol and process were designed and implemented with that purpose in mind. Third, we may have missed articles by deciding to not include the CINAHL and PsycINFO databases in addition to MEDLINE, Embase, and the instrument databases. Despite reviewing at least 50 articles most relevant to the search strategies for CINAHL and PsycINFO and finding none aligned with our inclusion criteria, it is possible there were articles relevant to our review in the rest of the search results from these two databases. Fourth, because we did not want to exclude articles from any region or language but are not fluent in all languages, we used Google Translate for some screening and review. It is, therefore, possible the translation provided did not allow us to sufficiently evaluate articles per our inclusion/exclusion criteria. For the two relevant articles not in English, we did complete data extraction with a fluent colleague.

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Overall, there may be additional limitations about which we are not aware that may have biased the results of our systematic review. Our hope is, however, we took steps to mitigate as many as possible.

CONCLUSION

Despite the novel, broad, and transdisciplinary approach to our systematic review, the current instrument landscape, limitations in the literature, and gaps in evidence on measure quality and clinical trial utility indicate there is a need to examine changes to the menstrual cycle in a more complete, inclusive, and standardized way. Rigorous formative research—across sociocultural contexts—that is focused on how all people who menstruate experience and understand their menstrual cycles and more fully addresses menstrual stigma can inform the development of new or modified instruments to meet this need. We also identified a need for greater evidence of the validity for existing and new instruments. For the clinical trial context, current draft FDA guidance on selecting, developing, or modifying fit-for-purpose PROs indicate there must be evidence to support the use of an instrument for the specific concepts of interest and context of use [37]. At a minimum, per this guidance, evidence would be needed to support the use of the instruments identified and assessed in this review in the clinical trial context with a broader patient population (i.e., context of use) and to measure the full scope of menstrual changes that people experience (i.e., concept of interest). In addition, the recent emergence of core outcome sets within areas like HMB and endometriosis will be useful to promote standardization of validated instruments, especially if these efforts are interdisciplinary and coordinated across research areas.

The findings of our review will be helpful in developing new or modified instruments that assess menstrual changes in a validated, comprehensive way. If used across the many fields that study menstrual health, data from these standardized instruments can contribute to an interdisciplinary, systemic, and holistic understanding of menstruation and the menstrual cycle. In turn, this improved understanding can be translated into ways to enhance the health and wellbeing of people who menstruate.

ACKNOWLEDGEMENTS

The authors would like to thank Laneta Dorflinger, Rebecca Callahan, and members of the Global Contraceptive Induced Menstrual Changes (CIMC) Task Force for their feedback on the draft review

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

protocol. We are very grateful for the guidance and assistance provided by the FHI 360 health sciences library throughout the development and implementation of our literature search strategy (Allison Burns and Carol Manion) and during retrieval of full text articles (Tamara Fasnacht). We would also like to acknowledge Betsy Costenbader, Kavita Nanda, and Global CIMC Task Force members for their feedback on the draft data extraction forms, and Kavita Nanda for clinical advice. We would also like to thank Valeria Bahamondes for her assistance in the full text review and data extraction of the two Portuguese papers that were included. The authors also appreciate Laneta Dorflinger and Kate McQueen for their review of and feedback on drafts of this manuscript.

SUPPORTING INFORMATION

S1 Table. Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) checklist

S2 Table. Search strategy for databases other than MEDLINE.

S3 File. Details on the original and updated searches

S4 Table. Fields of data extraction form.

S5 Table. All articles included after title/abstract screening and full text review.

S65 Table. Characteristics of sub-scales, items, general instruments.

REFERENCES

1. Radačić I, Bennoune K, Pūras D, Boly Barry K, Heller L, Šimonovic D, et al. Women's menstrual health should no longer be a taboo. In: United Nations Office of the High Commissioner of Human Rights [Internet]. 2019 [cited 24 Jul 2023]. Available: https://www.ohchr.org/en/news/2019/03/international-womens-day-8-march-2019

2. Critchley HOD, Babayev E, Bulun SE, Clark S, Garcia-Grau I, Gregersen PK, et al. Menstruation: science and society. Am J Obstet Gynecol. 2020;223: 624–664. doi:10.1016/j.ajog.2020.06.004

3. Hennegan J, Winkler IT, Bobel C, Keiser D, Hampton J, Larsson G, et al. Menstrual health: a definition for policy, practice, and research. Sex Reprod Health Matters. 2021;29: 1911618. doi:10.1080/26410397.2021.1911618

4. Gharib M. Why 2015 Was The Year Of The Period, And We Don't Mean Punctuation. National Public Radio. 31 Dec 2015.

5. ACOG Committee Opinion No. 651: Menstruation in Girls and Adolescents: Using the Menstrual Cycle as a Vital Sign. Obstetrics and gynecology. 2015;126: e143–e146. doi:10.1097/AOG.0000000000001215

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

6. Bobel C, Winkler IT, Fahs B, Hasson KA, Kissling EA, Roberts T-A, editors. The Palgrave Handbook of Critical Menstruation Studies. Singapore: Springer Singapore; 2020. doi:10.1007/978-981-15-0614-7

7. Wilson LC, Rademacher KH, Rosenbaum J, Callahan RL, Nanda G, Fry S, et al. Seeking synergies: understanding the evidence that links menstrual health and sexual and reproductive health and rights. Sex Reprod Health Matters. 2021;29: 1882791. doi:10.1080/26410397.2021.1882791

8. Guilló-Arakistain M. Challenging Menstrual Normativity: Nonessentialist Body Politics and Feminist Epistemologies of Health. In: Bobel C, Winkler I, Fahs B, Hasson K, Kissling EarT, editors. The Palgrave Handbook of Critical Menstruation Studies. Singapore: Springer Singapore; 2020. pp. 869–883. doi:10.1007/978-981-15-0614-7_63

9. Fraser IS, Critchley HOD, Munro MG, Broder M, Writing Group for this Menstrual Agreement Process. A process designed to lead to international agreement on terminologies and definitions used to describe abnormalities of menstrual bleeding. Fertil Steril. 2007;87: 466–76. doi:10.1016/j.fertnstert.2007.01.023

10. Fraser IS, Critchley HOD, Munro MG, Broder M. Can we achieve international agreement on terminologies and definitions used to describe abnormalities of menstrual bleeding? Hum Reprod. 2007;22: 635–43. doi:10.1093/humrep/del478

11. Munro MG, Critchley HOD, Broder MS, Fraser IS, FIGO Working Group on Menstrual Disorders. FIGO classification system (PALM-COEIN) for causes of abnormal uterine bleeding in nongravid women of reproductive age. Int J Gynaecol Obstet. 2011;113: 3–13. doi:10.1016/j.ijgo.2010.11.011

12. Munro MG, Critchley HOD, Fraser IS, FIGO Menstrual Disorders Committee. The two FIGO systems for normal and abnormal uterine bleeding symptoms and classification of causes of abnormal uterine bleeding in the reproductive years: 2018 revisions. Int J Gynaecol Obstet. 2018;143: 393–408. doi:10.1002/ijgo.12666

13. WHO/UNICEF. Consultation on draft long list of goal, target and indicator options for future global monitoring of water, sanitation and hygiene. [cited 24 Jul 2023]. Available: https://washdata.org/sites/default/files/documents/reports/2017-06/JMP-2012-post2015-consultation.pdf

14. Hoppes E, Nwachukwu C, Hennegan J, Blithe DL, Cordova-Gomez A, Critchley H, et al. Global research and learning agenda for building evidence on contraceptive-induced menstrual changes for research, product development, policies, and programs. Gates Open Res. 2022;6: 49. doi:10.12688/gatesopenres.13609.1

15. Belsey EM, Machin D, D'Arcangues C. The analysis of vaginal bleeding patterns induced by fertility regulating methods. World Health Organization Special Programme of Research, Development and Research Training in Human Reproduction. Contraception. 1986;34: 253–60. doi:10.1016/0010-7824(86)90006-5

16. Belsey EM, Farley TMM. The analysis of menstrual bleeding patterns: A review. Applied Stochastic Models and Data Analysis. 1987;3: 125–150. doi:10.1002/asm.3150030302

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

17. Mishell DR, Guillebaud J, Westhoff C, Nelson AL, Kaunitz AM, Trussell J, et al. Recommendations for standardization of data collection and analysis of bleeding in combined hormone contraceptive trials. Contraception. 2007;75: 11–15. doi:10.1016/j.contraception.2006.08.012

18. Creinin MD, Vieira CS, Westhoff CL, Mansour DJA. Recommendations for standardization of bleeding data analyses in contraceptive studies. Contraception. 2022;112: 14–22. doi:10.1016/j.contraception.2022.05.011

19. Bobel C. The Managed Body: Developing Girls and Menstrual Health in the Global South. Cham, Switzerland: Palgrave Macmillan, Springer Nature; 2019. doi:10.1007/978-3-319-89414-0

20. Hennegan J, Brooks DJ, Schwab KJ, Melendez-Torres GJ. Measurement in the study of menstrual health and hygiene: A systematic review and audit. PLoS One. 2020;15: e0232935. doi:10.1371/journal.pone.0232935

21. Matteson KA. Menstrual questionnaires for clinical and research use. Best Pract Res Clin Obstet Gynaecol. 2017;40: 44–54. doi:10.1016/j.bpobgyn.2016.09.009

22. Magnay JL, O'Brien S, Gerlinger C, Seitz C. A systematic review of methods to measure menstrual blood loss. BMC Womens Health. 2018;18: 142. doi:10.1186/s12905-018-0627-8

23. Magnay JL, O'Brien S, Gerlinger C, Seitz C. Pictorial methods to assess heavy menstrual bleeding in research and clinical practice: a systematic literature review. BMC Womens Health. 2020;20: 24. doi:10.1186/s12905-020-0887-y

24. Lee KMN, Junkins EJ, Luo C, Fatima UA, Cox ML, Clancy KBH. Investigating trends in those who experience menstrual bleeding changes after SARS-CoV-2 vaccination. Sci Adv. 2022;8: eabm7201. doi:10.1126/sciadv.abm7201

25. Edelman A, Boniface ER, Benhar E, Han L, Matteson KA, Favaro C, et al. Association Between Menstrual Cycle Length and Coronavirus Disease 2019 (COVID-19) Vaccination: A U.S. Cohort. Obstetrics and gynecology. 2022. doi:10.1097/AOG.0000000000004695

26. Gibson EA, Li H, Fruh V, Gabra M, Asokan G, Jukic AMZ, et al. Covid-19 vaccination and menstrual cycle length in the Apple Women's Health Study. NPJ Digit Med. 2022;5: 165. doi:10.1038/s41746-022-00711-9

27. Wesselink AK, Lovett SM, Weinberg J, Geller RJ, Wang TR, Regan AK, et al. COVID-19 vaccination and menstrual cycle characteristics: A prospective cohort study. Vaccine. 2023;41: 4327–4334. doi:10.1016/j.vaccine.2023.06.012

28. Kareem R, Sethi MR, Inayat S, Irfan M. The effect of COVID-19 vaccination on the menstrual pattern and mental health of the medical students: A mixed-methods study from a low and middle-income country. PLoS One. 2022;17: e0277288. doi:10.1371/journal.pone.0277288

29. Kim AM, Tingen CM, Woodruff TK. Sex bias in trials and treatment must end. Nature. 2010;465: 688–9. doi:10.1038/465688a

30. Beery AK, Zucker I. Sex bias in neuroscience and biomedical research. Neurosci Biobehav Rev. 2011;35: 565–72. doi:10.1016/j.neubiorev.2010.07.002

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

31. Shah K, McCormack CE, Bradbury NA. Do you know the sex of your cells? Am J Physiol Cell Physiol. 2014;306: C3-18. doi:10.1152/ajpcell.00281.2013

32. Woitowich NC, Beery A, Woodruff T. A 10-year follow-up study of sex inclusion in the biological sciences. Elife. 2020;9. doi:10.7554/eLife.56344

33. Sugimoto CR, Ahn Y-Y, Smith E, Macaluso B, Larivière V. Factors affecting sex-related reporting in medical research: a cross-disciplinary bibliometric analysis. Lancet. 2019;393: 550–559. doi:10.1016/S0140-6736(18)32995-7

34. Arnegard ME, Whitten LA, Hunter C, Clayton JA. Sex as a Biological Variable: A 5-Year Progress Report and Call to Action. J Womens Health (Larchmt). 2020. doi:10.1089/jwh.2019.8247

35. FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools) Resource. 2021 Edition. Silver Spring, MD & Bethesda, MD: US Food and Drug Administration & US National Institutes of Health; 2016. Available: https://www.ncbi.nlm.nih.gov/books/NBK338448/

36. US Food and Drug Administration. Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. 2009. Available: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-reported-outcome-measures-use-medical-product-development-support-labeling-claims

37. US Food and Drug Administration. Patient-Focused Drug Development: Selecting, Developing, or Modifying Fit-for-Purpose Clinical Outcome Assessments: Draft Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders. 2022. Available: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-focused-drug-development-selecting-developing-or-modifying-fit-purpose-clinical-outcome

38. US Food and Drug Administration. FDA Patient-Focused Drug Development Guidance Series for Enhancing the Incorporation of the Patient's Voice in Medical Product Development and Regulatory Decision Making. 6 Apr 2023 [cited 20 Sep 2023]. Available: https://www.fda.gov/drugs/development-approval-process-drugs/fda-patient-focused-drug-development-guidance-series-enhancing-incorporation-patients-voice-medical

39. US Food and Drug Administration. Principles for Selecting, Developing, Modifying, and Adapting Patient-Reported Outcome Instruments for Use in Medical Device Evaluation. 2022 Jan. Available: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/principles-selecting-developing-modifying-and-adapting-patient-reported-outcome-instruments-use

40. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. PLoS Med. 2021;18: e1003583. doi:10.1371/journal.pmed.1003583

41. Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ. 2021;372: n160. doi:10.1136/bmj.n160

42. Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, et al. PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. Syst Rev. 2021;10: 39. doi:10.1186/s13643-020-01542-z

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

43. Mackenzie A, Chung S, Hoppes E, Cartwright A, Mickler A. Measurement of changes to the menstrual cycle: A systematic review protocol. PROSPERO. 2023;CRD4202342. Available: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42023420358

44. Terwee CB, Jansma EP, Riphagen II, de Vet HCW. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res. 2009;18: 1115–23. doi:10.1007/s11136-009-9528-5

45. Veritas Health Innovation. Covidence Systematic Review Software. [cited 24 Jul 2023]. Available: https://www.covidence.org

46. US National Library of Medicine, US National Institutes of Health. NIH Common Data Elements Repository. [cited 11 Oct 2023]. Available: https://cde.nlm.nih.gov

47. University Library Vrije Universiteit Amsterdam. COSMIN database of systematic reviews of outcome measurement instruments. [cited 11 Oct 2023]. Available: https://database.cosmin.nl

48. Core Outcome Measures in Effectiveness Trials (COMET) Initiative. COMET database. [cited 11 Oct 2023]. Available: https://www.comet-initiative.org/Studies

49. Mapi Research Trust. ePROVIDE. [cited 11 Oct 2023]. Available: https://eprovide.mapi-trust.org

50. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine. Cambridge University Press; 2011. doi:10.1017/CBO9780511996214

51. Crossnohere NL, Brundage M, Calvert MJ, King M, Reeve BB, Thorner E, et al. International guidance on the selection of patient-reported outcome measures in clinical trials: a review. Qual Life Res. 2021;30: 21–40. doi:10.1007/s11136-020-02625-z

52. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res. 2010;19: 539–49. doi:10.1007/s11136-010-9606-8

53. Francis DO, McPheeters ML, Noud M, Penson DF, Feurer ID. Checklist to operationalize measurement characteristics of patient-reported outcome measures. Syst Rev. 2016;5: 129. doi:10.1186/s13643-016-0307-4

54. Rothman M, Burke L, Erickson P, Leidy NK, Patrick DL, Petrie CD. Use of existing patient-reported outcome (PRO) instruments and their modification: the ISPOR Good Research Practices for Evaluating and Documenting Content Validity for the Use of Existing Instruments and Their Modification PRO Task Force Report. Value Health. 2009;12: 1075–83. doi:10.1111/j.1524-4733.2009.00603.x

55. Patrick DL, Burke LB, Gwaltney CJ, Leidy NK, Martin ML, Molsen E, et al. Content validity--establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: part 2--assessing respondent understanding. Value Health. 2011;14: 978–88. doi:10.1016/j.jval.2011.06.013

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

56. Snyder C, Crossnohere N, King M, Reeve BB, Bottomley A, Calvert M, et al. The PROTEUS-Trials Consortium: Optimizing the use of patient-reported outcomes in clinical trials. Clin Trials. 2022;19: 277–284. doi:10.1177/17407745221077691

57. Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. Qual Life Res. 2013;22: 1889–905. doi:10.1007/s11136-012-0344-y

58. Johnston-Robledo I, Chrisler JC. The Menstrual Mark: Menstruation as Social Stigma. In: Bobel C, Winkler I, Fahs B, Hasson K, Kissling E, Roberts T, editors. The Palgrave Handbook of Critical Menstruation Studies. Singapore: Springer Singapore; 2020. pp. 181–199. doi:10.1007/978-981-15-0614-7_17

59. Erci B, Güngörmüş Z, Oztürk S. Psychometric validation of the Women's Health Questionnaire in menopausal women. Health Care Women Int. 2014;35: 566–79. doi:10.1080/07399332.2013.841698

60. Tatlock S, Abraham L, Bushmakin A, Moffatt M, Williamson N, Coon C, et al. Psychometric evaluation of electronic diaries assessing side-effects of hormone therapy. Climacteric. 2018;21: 594–600. doi:10.1080/13697137.2018.1517738

61. Lamping DL, Rowe P, Clarke A, Black N, Lessof L. Development and validation of the Menorrhagia Outcomes Questionnaire. Br J Obstet Gynaecol. 1998;105: 766–79. doi:10.1111/j.1471-0528.1998.tb10209.x

62. Jenkinson C, Peto V, Coulter A. Measuring change over time: a comparison of results from a global single item of health status and the multi-dimensional SF-36 health status survey questionnaire in patients presenting with menorrhagia. Qual Life Res. 1994;3: 317–21. doi:10.1007/BF00451723

63. Cooper NAM, Rivas C, Munro MG, Critchley HOD, Clark TJ, Matteson KA, et al. Standardising outcome reporting for clinical trials of interventions for heavy menstrual bleeding: Development of a core outcome set. BJOG. 2023. doi:10.1111/1471-0528.17473

64. Duffy J, Hirsch M, Vercoe M, Abbott J, Barker C, Collura B, et al. A core outcome set for future endometriosis research: an international consensus development study. BJOG. 2020;127: 967–974. doi:10.1111/1471-0528.16157

65. Nogueira-Silva C, Costa P, Martins C, Barata S, Alho C, Calhaz-Jorge C, et al. Validação da Versão Portuguesa do Questionário EHP-30 (The Endometriosis Health Profile-30). Acta Med Port. 2015;28: 347–356. doi:10.20344/amp.5778

66. Mengarda CV, Passos EP, Picon P, Costa AF, Picon PD. Validação de versão para o português de questionário sobre qualidade de vida para mulher com endometriose (Endometriosis Health Profile Questionnaire - EHP-30). Revista Brasileira de Ginecologia e Obstetrícia. 2008;30: 384–392. doi:10.1590/S0100-72032008000800003

67. Schumacher U, Schumacher J, Mellinger U, Gerlinger C, Wienke A, Endrikat J. Estimation of menstrual blood loss volume based on menstrual diary and laboratory data. BMC Womens Health. 2012;12: 24. doi:10.1186/1472-6874-12-24

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

68. Weller A, Weller L. Assessment of menstrual regularity and irregularity using self-reports and objective criteria. Journal of Psychosomatic Obstetrics & Gynecology. 1998;19: 111–116. doi:10.3109/01674829809048504

69. Paramsothy P, Harlow SD, Elliott MR, Lisabeth LD, Crawford SL, Randolph JF. Classifying menopause stage by menstrual calendars and annual interviews: need for improved questionnaires. Menopause. 2013;20: 727–35. doi:10.1097/GME.0b013e3182825ff2

70. Johannes CB, Crawford SL, Woods J, Goldstein RB, Tran D, Mehrotra S, et al. An Electronic Menstrual Cycle Calendar: Comparison of Data Quality With a Paper Version. Menopause. 2000;7: 200–208. doi:10.1097/00042192-200007030-00011

71. Toxqui L, Pérez-Granados AM, Blanco-Rojo R, Wright I, Vaquero MP. A simple and feasible questionnaire to estimate menstrual blood loss: relationship with hematological and gynecological parameters in young women. BMC Womens Health. 2014;14: 71. doi:10.1186/1472-6874-14-71

72. Calaf J, Palacios S, Cristóbal I, Cañete ML, Monleón J, Fernández J, et al. Validation of the Spanish version of the Uterine Fibroid Symptom and Quality of Life (UFS-QoL) questionnaire in women with uterine myomatosis. Med Clin (Barc). 2020;154: 207–213. doi:10.1016/j.medcli.2019.05.027

73. Matteson KA, Scott DM, Raker CA, Clark MA. The menstrual bleeding questionnaire: development and validation of a comprehensive patient-reported outcome instrument for heavy menstrual bleeding. BJOG. 2015;122: 681–9. doi:10.1111/1471-0528.13273

74. Lancastle D, Kopp Kallner H, Hale G, Wood B, Ashcroft L, Driscoll H. Development of a brief menstrual quality of life measure for women with heavy menstrual bleeding. BMC Womens Health. 2023;23. doi:10.1186/s12905-023-02235-0

75. Nguyen AM, Arbuckle R, Korver T, Chen F, Taylor B, Turnbull A, et al. Psychometric validation of the dysmenorrhea daily diary (DysDD): a patient-reported outcome for dysmenorrhea. Quality of Life Research. 2017;26: 2041–2055. doi:10.1007/s11136-017-1562-0

76. Small CM, Manatunga AK, Marcus M. Validity of Self-Reported Menstrual Cycle Length. Ann Epidemiol. 2007;17: 163–170. doi:10.1016/j.annepidem.2006.05.005

77. Jukic AMZ, Weinberg CR, Wilcox AJ, McConnaughey DR, Hornsby P, Baird DD. Accuracy of Reporting of Menstrual Cycle Length. Am J Epidemiol. 2007;167: 25–33. doi:10.1093/aje/kwm265

78. Wegienka G, Baird DD. A comparison of recalled date of last menstrual period with prospectively recorded dates. J Womens Health (Larchmt). 2005;14: 248–52. doi:10.1089/jwh.2005.14.248

79. Bachand AM, Cragin LA, Reif JS. Reliability of Retrospectively Assessed Categorical Menstrual Cycle Length Data. Ann Epidemiol. 2009;19: 501–503. doi:10.1016/j.annepidem.2009.03.015

80. Ruta DA, Garratt AM, Chadha YC, Flett GM, Hall MH, Russell IT. Assessment of patients with menorrhagia: How valid is a structured clinical history as a measure of health status? Quality of Life Research. 1995;4: 33–40. doi:10.1007/BF00434381

81. Vitonis AF, Vincent K, Rahmioglu N, Fassbender A, Buck Louis GM, Hummelshoj L, et al. World Endometriosis Research Foundation Endometriosis Phenome and biobanking harmonization

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

project: II. Clinical and covariate phenotype data collection in endometriosis research. Fertil Steril. 2014;102: 1223–1232. doi:10.1016/j.fertnstert.2014.07.1244

82. Shin H, Park Y-J, Cho I. Development and psychometric validation of the Menstrual Health Instrument (MHI) for adolescents in Korea. Health Care Women Int. 2018;39: 1090–1109. doi:10.1080/07399332.2017.1423487

83. Gray TG, Moores KL, James E, Connor ME, Jones GL, Radley SC. Development and initial validation of an electronic personal assessment questionnaire for menstrual, pelvic pain and gynaecological hormonal disorders (ePAQ-MPH). European Journal of Obstetrics & Gynecology and Reproductive Biology. 2019;238: 148–156. doi:10.1016/j.ejogrb.2019.05.024

84. Cho H-H, Yoon Y-S. Development of an endometriosis self-assessment tool for patient. Obstet Gynecol Sci. 2022;65: 256–265. doi:10.5468/ogs.21252

85. Caruso BA, Portela G, McManus S, Clasen T. Assessing Women's Menstruation Concerns and Experiences in Rural India: Development and Validation of a Menstrual Insecurity Measure. Int J Environ Res Public Health. 2020;17: 3468. doi:10.3390/ijerph17103468

86. Kantarovich D, Dillane KE, Garrison EF, Oladosu FA, Schroer MS, Roth GE, et al. Development and validation of a real-time method characterizing spontaneous pain in women with dysmenorrhea. Journal of Obstetrics and Gynaecology Research. 2021;47: 1472–1480. doi:10.1111/jog.14663

87. Rodrigues JC, Avila MA, dos Reis FJJ, Carlessi RM, Godoy AG, Arruda GT, et al. 'Painting my pain': the use of pain drawings to assess multisite pain in women with primary dysmenorrhea. BMC Womens Health. 2022;22. doi:10.1186/s12905-022-01945-1

88. Jukic AMZ, Weinberg CR, Baird DD, Hornsby PP, Wilcox AJ. Measuring Menstrual Discomfort. Epidemiology. 2008;19: 846–850. doi:10.1097/EDE.0b013e318187ac9e

89. Deal LS, DiBenedetti D, Williams VS, Fehnel SE. The development and validation of the daily electronic Endometriosis Pain and Bleeding Diary. Health Qual Life Outcomes. 2010;8: 64. doi:10.1186/1477-7525-8-64

90. Farquhar CM, Roberts H, Okonkwo QL, Stewart AW. A pilot survey of the impact of menstrual cycles on adolescent health. Aust N Z J Obstet Gynaecol. 2009;49: 531–6. doi:10.1111/j.1479-828X.2009.01062.x

91. Darabi F, Yaseri M, Rohban A, Khalajabadi-Farahani F. Development and Psychometric Properties of Menstrual Health Seeking Behaviors Questionnaire (MHSBQ-42) in Female Adolescents. J Reprod Infertil. 2018;19: 229–236.

92. Garg S, Marimuthu Y, Bhatnagar N, Singh MmC, Borle A, Basu S, et al. Development and validation of a menstruation-related activity restriction questionnaire among adolescent girls in urban resettlement colonies of Delhi. Indian Journal of Community Medicine. 2021;46: 57. doi:10.4103/ijcm.IJCM_183_20

93. Ramaiya A, Sood S. What are the psychometric properties of a menstrual hygiene management scale: a community-based cross-sectional study. BMC Public Health. 2020;20: 525. doi:10.1186/s12889-020-08627-3

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

94. Hennegan J, Nansubuga A, Smith C, Redshaw M, Akullo A, Schwab KJ. Measuring menstrual hygiene experience: development and validation of the Menstrual Practice Needs Scale (MPNS-36) in Soroti, Uganda. BMJ Open. 2020;10: e034461. doi:10.1136/bmjopen-2019-034461

95. Das M. Socio-cultural aspects of menstruation: An anthropological purview. East Anthropol. 2008;61: 227–240.

96. Tan DA, Haththotuwa R, Fraser IS. Cultural aspects and mythologies around menstruation and abnormal uterine bleeding. Best Pract Res Clin Obstet Gynaecol. 2016; 1–13. doi:10.1016/j.bpobgyn.2016.09.015

97. Clancy K. Period: The Real Story of Menstruation. Princeton, New Jersey, US: Princeton University Press; 2023.

98. Lee K, Junkins E, Fatima U, Clancy K. Measuring menstruation: methodological difficulties in studying things we don't talk about. American Journal of Human Biology. 2022;34: 76.1. doi:10.1002/ajhb.23740

99. Bull JR, Rowland SP, Scherwitzl EB, Scherwitzl R, Danielsson KG, Harper J. Real-world menstrual cycle characteristics of more than 600,000 menstrual cycles. NPJ Digit Med. 2019;2: 83. doi:10.1038/s41746-019-0152-7

100. Li H, Gibson EA, Jukic AMZ, Baird DD, Wilcox AJ, Curry CL, et al. Menstrual cycle length variation by demographic characteristics from the Apple Women's Health Study. NPJ Digit Med. 2023;6: 100. doi:10.1038/s41746-023-00848-1

101. Li K, Urteaga I, Wiggins CH, Druet A, Shea A, Vitzthum VJ, et al. Characterizing physiological and symptomatic variation in menstrual cycles using self-tracked mobile-health data. NPJ Digit Med. 2020;3: 79. doi:10.1038/s41746-020-0269-8

102. Wiederhold BK. Femtech: Digital Help for Women's Health Care Across the Life Span. Cyberpsychol Behav Soc Netw. 2021;24: 697–698. doi:10.1089/cyber.2021.29230.editorial

103. Krishnamurti T, Birru Talabi M, Callegari LS, Kazmerski TM, Borrero S. A Framework for Femtech: Guiding Principles for Developing Digital Reproductive Health Tools in the United States. J Med Internet Res. 2022;24: e36338. doi:10.2196/36338

104. Eremenco S, Coons SJ, Paty J, Coyne K, Bennett A V, McEntegart D, et al. PRO data collection in clinical trials using mixed modes: report of the ISPOR PRO mixed modes good research practices task force. Value Health. 2014;17: 501–16. doi:10.1016/j.jval.2014.06.005

105. Harlow SD, Lin X, Ho MJ. Analysis of menstrual diary data across the reproductive life span applicability of the bipartite model approach and the importance of within-woman variance. J Clin Epidemiol. 2000;53: 722–33. doi:10.1016/s0895-4356(99)00202-4

106. Harlow SD, Paramsothy P. Menstruation and the menopausal transition. Obstet Gynecol Clin North Am. 2011;38: 595–607. doi:10.1016/j.ogc.2011.05.010

107. US National Institutes of Health. NIH Strategic Plan to Advance Research on the Health and Well-being of Sexual and Gender Minorities FYs 2021-2025. 2020 [cited 24 Jul 2023]. Available: https://dpcpsi.nih.gov/sites/default/files/SGMStrategicPlan_2021_2025.pdf

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

108. van Eijkeren MA, Scholten PC, Christiaens GC, Alsbach GP, Haspels AA. The alkaline hematin method for measuring menstrual blood loss--a modification and its clinical use in menorrhagia. Eur J Obstet Gynecol Reprod Biol. 1986;22: 345–51. doi:10.1016/0028-2243(86)90124-3

109. Mansfield PK, Voda A, Allison G. Validating a pencil-and-paper measure of perimenopausal menstrual blood loss. Women's Health Issues. 2004;14: 242–247. doi:10.1016/j.whi.2004.07.005

110. Chimbira TH, Anderson AB, Turnbull A c. Relation between measured menstrual blood loss and patient's subjective assessment of loss, duration of bleeding, number of sanitary towels used, uterine weight and endometrial surface area. Br J Obstet Gynaecol. 1980;87: 603–9. doi:10.1111/j.1471-0528.1980.tb05013.x

111. Fraser IS, Warner P, Marantos PA. Estimating menstrual blood loss in women with normal and excessive menstrual fluid volume. Obstetrics and gynecology. 2001;98: 806–14. doi:10.1016/s0029-7844(01)01581-2

112. Gannon MJ, Day P, Hammadieh N, Johnson N. A new method for measuring menstrual blood loss and its use in screening women before endometrial ablation. BJOG. 1996;103: 1029–1033. doi:10.1111/j.1471-0528.1996.tb09556.x

113. Gleeson N, Devitt M, Buggy F, Bonnar J. Menstrual Blood Loss Measurement with Gynaeseal. Australian and New Zealand Journal of Obstetrics and Gynaecology. 1993;33: 79–80. doi:10.1111/j.1479-828X.1993.tb02061.x

114. Gudmundsdottir BR, Hjaltalin EF, Bragadottir G, Hauksson A, Geirsson RT, Onundarson PT. Quantification of menstrual flow by weighing protective pads in women with normal, decreased or increased menstruation. Acta Obstet Gynecol Scand. 2009;88: 275–9. doi:10.1080/00016340802673162

115. Heath AL, Skeaff CM, Gibson RS. Validation of a questionnaire method for estimating extent of menstrual blood loss in young adult women. J Trace Elem Med Biol. 1999;12: 231–5. doi:10.1016/S0946-672X(99)80063-7

116. Barr F, Brabin L, Agbaje O. A pictorial chart for managing common menstrual disorders in Nigerian adolescents. Int J Gynaecol Obstet. 1999;66: 51–3. doi:10.1016/s0020-7292(99)00025-9

117. Haberland C, Filonenko A, Seitz C, Börner M, Gerlinger C, Doll H, et al. Validation of a menstrual pictogram and a daily bleeding diary for assessment of uterine fibroid treatment efficacy in clinical studies. J Patient Rep Outcomes. 2020;4: 97. doi:10.1186/s41687-020-00263-0

118. Hald K, Lieng M. Assessment of Periodic Blood Loss: Interindividual and Intraindividual Variations of Pictorial Blood Loss Assessment Chart Registrations. J Minim Invasive Gynecol. 2014;21: 662–668. doi:10.1016/j.jmig.2014.01.015

119. Higham JM, O'Brien PM, Shaw RW. Assessment of menstrual blood loss using a pictorial chart. Br J Obstet Gynaecol. 1990;97: 734–9. doi:10.1111/j.1471-0528.1990.tb16249.x

120. Janssen CA, Scholten PC, Heintz AP. A simple visual assessment technique to discriminate between menorrhagia and normal menstrual blood loss. Obstetrics and gynecology. 1995;85: 977–82. doi:10.1016/0029-7844(95)00062-V

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

121. Larsen L, Coyne K, Chwalisz K. Validation of the menstrual pictogram in women with leiomyomata associated with heavy menstrual bleeding. Reprod Sci. 2013;20: 680–7. doi:10.1177/1933719112463252

122. Magnay JL, Nevatte TM, Seitz C, O'Brien S. A new menstrual pictogram for use with feminine products that contain superabsorbent polymers. Fertil Steril. 2013;100: 1715-21.e1–4. doi:10.1016/j.fertnstert.2013.08.028

123. Magnay JL, Nevatte TM, O'Brien S, Gerlinger C, Seitz C. Validation of a new menstrual pictogram (superabsorbent polymer-c version) for use with ultraslim towels that contain superabsorbent polymers. Fertil Steril. 2014;101: 515–22. doi:10.1016/j.fertnstert.2013.10.051

124. Reid PC, Coker A, Coltart R. Assessment of menstrual blood loss using a pictorial chart: a validation study. BJOG. 2000;107: 320–322. doi:10.1111/j.1471-0528.2000.tb13225.x

125. Sanchez J, Andrabi S, Bercaw JL, Dietrich JE. Quantifying the PBAC in a Pediatric and Adolescent Gynecology Population. Pediatr Hematol Oncol. 2012;29: 479–484. doi:10.3109/08880018.2012.699165

126. Wyatt KM, Dimmock PW, Walker TJ, O'Brien PMS. Determination of total menstrual blood loss. Fertil Steril. 2001;76: 125–131. doi:10.1016/S0015-0282(01)01847-7

127. de Arruda GT, Driusso P, Rodrigues JC, de Godoy AG, Avila MA. Numerical rating scale for dysmenorrhea-related pain: a clinimetric study. Gynecological Endocrinology. 2022;38. doi:10.1080/09513590.2022.2099831

128. Pokrzywinski RM, Soliman AM, Snabes MC, Chen J, Taylor HS, Coyne KS. Responsiveness and thresholds for clinically meaningful changes in worst pain numerical rating scale for dysmenorrhea and nonmenstrual pelvic pain in women with moderate to severe endometriosis. Fertil Steril. 2021;115: 423–430. doi:10.1016/j.fertnstert.2020.07.013

129. Gerlinger C, Schumacher U, Faustmann T, Colligs A, Schmitz H, Seitz C. Defining a minimal clinically important difference for endometriosis-associated pelvic pain measured on a visual analog scale: analyses of two placebo-controlled, randomized trials. Health Qual Life Outcomes. 2010;8: 138. doi:10.1186/1477-7525-8-138

130. Gerlinger C, Schumacher U, Wentzeck R, Uhl-Hochgräber K, Solomayer EF, Schmitz H, et al. How can we measure endometriosis-associated pelvic pain? J Endometr. 2012;4: 109–116. doi:10.5301/JE.2012.9725

131. Ching-Hsing H, Meei-Ling G, Hsin-Chun M, Chung-Yi L. The development and psychometric testing of a self-care scale for dysmenorrhic adolescents. The journal of nursing research. 2004;12: 119–30. doi:10.1097/01.jnr.0000387495.01557.aa

132. Wong CL, Ip WY, Choi KC, Shiu TY. Translation and validation of the Chinese-Cantonese version of the Adolescent Dysmenorrhic Self-Care Scale in Hong Kong adolescent girls. J Clin Nurs. 2013;22: 1510–1520. doi:10.1111/jocn.12019

133. Chen CX, Murphy T, Ofner S, Yahng L, Krombach P, LaPradd M, et al. Development and Testing of the Dysmenorrhea Symptom Interference (DSI) Scale. West J Nurs Res. 2021;43: 364–373. doi:10.1177/0193945920942252

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

134. Chauvet P, Auclair C, Mourgues C, Canis M, Gerbaud L, Bourdel N. Psychometric properties of the French version of the Endometriosis Health Profile-30, a health-related quality of life instrument. J Gynecol Obstet Hum Reprod. 2017;46: 235–242. doi:10.1016/j.jogoh.2017.02.004

135. Grundström H, Rauden A, Wikman P, Olovsson M. Psychometric evaluation of the Swedish version of the 30-item endometriosis health profile (EHP-30). BMC Womens Health. 2020;20: 204. doi:10.1186/s12905-020-01067-6

136. Grundström H, Rauden A, Olovsson M. Cross-cultural adaptation of the Swedish version of Endometriosis Health Profile-30. J Obstet Gynaecol (Lahore). 2020;40: 969–973. doi:10.1080/01443615.2019.1676215

137. Hansen KE, Lambek R, Røssaak K, Egekvist AG, Marschall H, Forman A, et al. Health-related quality of life in women with endometriosis: psychometric validation of the Endometriosis Health Profile 30 questionnaire using confirmatory factor analysis. Hum Reprod Open. 2022;2022. doi:10.1093/hropen/hoab042

138. Jenkinson C, Kennedy S, Jones G. Evaluation of the American version of the 30-item Endometriosis Health Profile (EHP-30). Quality of Life Research. 2008;17: 1147–1152. doi:10.1007/s11136-008-9403-9

139. Jia S-Z, Leng J-H, Sun P-R, Lang J-H. Translation and psychometric evaluation of the simplified Chinese-version Endometriosis Health Profile-30. Human Reproduction. 2013;28: 691–697. doi:10.1093/humrep/des426

140. Jones G, Kennedy S, Barnard A, Wong J, Jenkinson C. Development of an endometriosis quality-of-life instrument: The Endometriosis Health Profile-30. Obstetrics and gynecology. 2001;98: 258–64. doi:10.1016/s0029-7844(01)01433-8

141. Jones G, Jenkinson C, Kennedy S. Evaluating the responsiveness of the endometriosis health profile questionnaire: The EHP-30. Quality of Life Research. 2004;13: 705–713. doi:10.1023/B:QURE.0000021316.79349.af

142. Jones G, Jenkinson C, Taylor N, Mills A, Kennedy S. Measuring quality of life in women with endometriosis: tests of data quality, score reliability, response rate and scaling assumptions of the Endometriosis Health Profile Questionnaire. Human Reproduction. 2006;21: 2686–2693. doi:10.1093/humrep/del231

143. Khong S-Y, Lam A, Luscombe G. Is the 30-item Endometriosis Health Profile (EHP-30) suitable as a self-report health status instrument for clinical trials? Fertil Steril. 2010;94: 1928–1932. doi:10.1016/j.fertnstert.2010.01.047

144. Maiorana A, Scafidi Fonti GM, Audino P, Rosini R, Alio L, Oliveri AM, et al. The role of EHP-30 as specific instrument to assess the quality of life of Italian women with endometriosis. Minerva Ginecol. 2012;64: 231–8. Available: http://www.ncbi.nlm.nih.gov/pubmed/22635018

145. Mansor M, Chong MC, Chui PL, Hamdan M, Basha MAMK. The psychometric properties test of the Malay version of the endometriosis health profile-30. Saudi Med J. 2023;44. doi:10.15537/smj.2023.44.9.20230228

146. Marí-Alexandre J, García-Oms J, Agababyan C, Belda-Montesinos R, Royo-Bolea S, Varo-Gómez B, et al. Toward an improved assessment of quality of life in endometriosis: evaluation of the

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Spanish version of the Endometriosis Health Profile 30. Journal of Psychosomatic Obstetrics and Gynecology. 2022;43. doi:10.1080/0167482X.2020.1795827

147. Nojomi M, Bijari B, Akhbari R, Kashanian M. The Assessment of Reliability and Validity of Persian Version of the Endometriosis Health Profile (EHP-30). Iran J Med Sci. 2011;36: 84–9. Available: http://www.ncbi.nlm.nih.gov/pubmed/23358588

148. van de Burgt TJM, Hendriks JCM, Kluivers KB. Quality of life in endometriosis: evaluation of the Dutch-version Endometriosis Health Profile–30 (EHP-30). Fertil Steril. 2011;95: 1863–1865. doi:10.1016/j.fertnstert.2010.11.009

149. van de Burgt TJM, Kluivers KB, Hendriks JCM. Responsiveness of the Dutch Endometriosis Health Profile-30 (EHP-30) questionnaire. European Journal of Obstetrics & Gynecology and Reproductive Biology. 2013;168: 92–94. doi:10.1016/j.ejogrb.2012.12.037

150. Verket NJ, Andersen MH, Sandvik L, Tanbo TG, Qvigstad E. Lack of cross-cultural validity of the Endometriosis Health Profile-30. J Endometr Pelvic Pain Disord. 2018;10: 107–115. doi:10.1177/2284026518780638

151. Wickström KW, Spira J, Edelstam G. Responsiveness of the Endometriosis Health Profile-30 questionnaire in a Swedish sample: an observational study. Clin Exp Obstet Gynecol. 2017;44: 413–418. doi:10.12891/ceog3599.2017

152. Aubry G, Panel P, Thiollier G, Huchon C, Fauconnier A. Measuring health-related quality of life in women with endometriosis: comparing the clinimetric properties of the Endometriosis Health Profile-5 (EHP-5) and the EuroQol-5D (EQ-5D). Human Reproduction. 2017;32: 1258–1269. doi:10.1093/humrep/dex057

153. Fauconnier A, Huchon C, Chaillou L, Aubry G, Renouvel F, Panel P. Development of a French version of the Endometriosis Health Profile 5 (EHP-5): cross-cultural adaptation and psychometric evaluation. Quality of Life Research. 2017;26: 213–220. doi:10.1007/s11136-016-1346-y

154. Mikuš M, Matak L, Vujić G, Škegro B, Škegro I, Augustin G, et al. The short form endometriosis health profile questionnaire (EHP-5): psychometric validity assessment of a Croatian version. Arch Gynecol Obstet. 2023;307: 87–92. doi:10.1007/s00404-022-06691-1

155. Selcuk S, Sahin S, Demirci O, Aksoy B, Eroglu M, Ay P, et al. Translation and validation of the Endometriosis Health Profile (EHP-5) in patients with laparoscopically diagnosed endometriosis. European Journal of Obstetrics & Gynecology and Reproductive Biology. 2015;185: 41–44. doi:10.1016/j.ejogrb.2014.11.039

156. Gater A, Taylor F, Seitz C, Gerlinger C, Wichmann K, Haberland C. Development and content validation of two new patient-reported outcome measures for endometriosis: the Endometriosis Symptom Diary (ESD) and Endometriosis Impact Scale (EIS). J Patient Rep Outcomes. 2020;4: 13. doi:10.1186/s41687-020-0177-3

157. Deal LS, Williams VSL, DiBenedetti DB, Fehnel SE. Development and psychometric evaluation of the Endometriosis Treatment Satisfaction Questionnaire. Quality of Life Research. 2010;19: 899–905. doi:10.1007/s11136-010-9640-6

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

158. Li L, Huangfu L, Chai H, He W, Song H, Zou X, et al. Development of a functional and emotional measure of dysmenorrhea (FEMD) in Chinese university women. Health Care Women Int. 2012;33: 97–108. doi:10.1080/07399332.2011.603863

159. Yamada K, Adachi T, Kubota Y, Takeda T, Iseki M. Developing a Japanese version of the Injustice Experience Questionnaire-chronic and the contribution of perceived injustice to severity of menstrual pain: a web-based cross-sectional study. Biopsychosoc Med. 2019;13: 17. doi:10.1186/s13030-019-0158-z

160. Habiba M, Julian S, Taub N, Clark M, Rashid A, Baker R, et al. Limited role of multi-attribute utility scale and SF-36 in predicting management outcome of heavy menstrual bleeding. European Journal of Obstetrics & Gynecology and Reproductive Biology. 2010;148: 81–85. doi:10.1016/j.ejogrb.2009.09.021

161. Pattison H, Daniels J, Kai J, Gupta J. The measurement properties of the menorrhagia multi-attribute quality-of-life scale: a psychometric analysis. BJOG. 2011;118: 1528–1531. doi:10.1111/j.1471-0528.2011.03057.x

162. Shaw RW, Brickley MR, Evans L, Edwards MJ. Perceptions of women on the impact of menorrhagia on their health using multi-attribute utility assessment. BJOG. 1998;105: 1155–1159. doi:10.1111/j.1471-0528.1998.tb09968.x

163. Bargiota S, Bonotis K, Garyfallos G, Messinis I, Angelopoulos N. The Psychometric Properties of Menstrual Attitudes Questionnaire: A validity Study in Greek Women. Int J Innov Res Sci Eng Technol. 2016;5: 1754–1765. doi:10.15680/IJIRSET.2016.0502109

164. Bramwell RS, Biswas EL, Anderson C. Using the Menstrual Attitude Questionnaire with a British and an Indian sample. J Reprod Infant Psychol. 2002;20: 159–170. doi:10.1080/026468302760270818

165. Brooks-Gunn J, Ruble DN. The menstrual attitude questionnaire. Psychosom Med. 1980;42: 503–12. doi:10.1097/00006842-198009000-00005

166. Firat MZ, Kulakaç Ö, Öncel S, Akcan A. Menstrual Attitude Questionnaire: confirmatory and exploratory factor analysis with Turkish samples. J Adv Nurs. 2009;65: 652–662. doi:10.1111/j.1365-2648.2008.04919.x

167. Kawata R, Endo M, Rai SK, Ohashi K. Development of a scale to evaluate negative menstrual attitudes among Nepalese women. Reprod Health. 2022;19: 120. doi:10.1186/s12978-022-01426-6

168. Darabi F, Yaseri M, Rohban A, Khalajabadi-Farahani F. Development and Psychometric Properties of Menstrual Health Seeking Behaviors Questionnaire (MHSBQ-42) in Female Adolescents. J Reprod Infertil. 2018;19: 229–236. Available: http://www.ncbi.nlm.nih.gov/pubmed/30746338

169. Aubeeluck A, Maguire M. The Menstrual Joy Questionnaire Items Alone Can Positively Prime Reporting of Menstrual Attitudes and Symptoms. Psychol Women Q. 2002;26: 160–162. doi:10.1111/1471-6402.00054

170. Hennegan J, Nansubuga A, Akullo A, Smith C, Schwab KJ. The Menstrual Practices Questionnaire (MPQ): development, elaboration, and implications for future research. Glob Health Action. 2020;13: 1829402. doi:10.1080/16549716.2020.1829402

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

171. Roberts T-A. Female Trouble: The Menstrual Self-Evaluation Scale and Women's Self-Objectification. Psychol Women Q. 2004;28: 22–26. doi:10.1111/j.1471-6402.2004.00119.x

172. Trego LL. Development of the Military Women's Attitudes Toward Menstrual Suppression Scale: From Construct Definition to Pilot Testing. J Nurs Meas. 2009;17: 45–72. doi:10.1891/1061-3749.17.1.45

173. Coyne KS, Soliman AM, Margolis MK, Thompson CL, Chwalisz K. Validation of the 4 week recall version of the Uterine Fibroid Symptom and Health-related Quality of Life (UFS-QOL) Questionnaire. Curr Med Res Opin. 2017;33: 193–200. doi:10.1080/03007995.2016.1248382

174. Coyne KS, Harrington A, Currie BM, Chen J, Gillard P, Spies JB. Psychometric validation of the 1-month recall Uterine Fibroid Symptom and Health-Related Quality of Life questionnaire (UFS-QOL). J Patient Rep Outcomes. 2019;3: 57. doi:10.1186/s41687-019-0146-x

175. Harding G, Coyne KS, Thompson CL, Spies JB. The responsiveness of the uterine fibroid symptom and health-related quality of life questionnaire (UFS-QOL). Health Qual Life Outcomes. 2008;6: 99. doi:10.1186/1477-7525-6-99

176. Keizer AL, van Kesteren PJM, Terwee C, de Lange ME, Hehenkamp WJK, Kok HS. Uterine Fibroid Symptom and Quality of Life questionnaire (UFS-QOL NL) in the Dutch population: a validation study. BMJ Open. 2021;11: e052664. doi:10.1136/bmjopen-2021-052664

177. Oliveira Brito LG, Malzone-Lott DA, Sandoval Fagundes MF, Magnani PS, Fernandes Arouca MA, Poli-Neto OB, et al. Translation and validation of the Uterine Fibroid Symptom and Quality of Life (UFS-QOL) questionnaire for the Brazilian Portuguese language. Sao Paulo Medical Journal. 2017;135: 107–115. doi:10.1590/1516-3180.2016.0223281016

178. Silva R, Gomes M, Castro R, Bonduki C, Girão M. Uterine Fibroid Symptom - Quality of Life questionnaire translation and validation into Brazilian Portuguese. Revista Brasileira de Ginecologia e Obstetrícia / RBGO Gynecology and Obstetrics. 2016;38: 518–523. doi:10.1055/s-0036-1593833

179. Spies JB, Coyne K, Guaou Guaou N, Boyle D, Skyrnarz-Murphy K, Gonzalves SM. The UFS-QOL, a new disease-specific symptom and health-related quality of life questionnaire for leiomyomata. Obstetrics and gynecology. 2002;99: 290–300. doi:10.1016/s0029-7844(01)01702-1

180. Yeung S, Kwok JW, Law S, Chung JP, Chan SS. Uterine Fibroid Symptom and Health-related Quality of Life Questionnaire: a Chinese translation and validation study. Hong Kong Medical Journal. 2019;25: 453–459. doi:10.12809/hkmj198064

181. Yu S-C, Hsu H-P, Guo J-L, Chen S-F, Huang S-H, Chen Y-C, et al. Exploration of the experiences of working stressors and coping strategies associated with menstrual symptoms among nurses with shifting schedules: a Q methodology investigation. BMC Nurs. 2021;20: 238. doi:10.1186/s12912-021-00759-0

182. de Arruda GT, de Melo Mantovan SG, Da Roza T, Silva BI da, Tonon da Luz SC, Avila MA. WHODAS measurement properties for women with dysmenorrhea. Health Qual Life Outcomes. 2023;21. doi:10.1186/s12955-023-02140-y

183. Abu-Rafea BF, Vilos GA, Al Jasser RS, Al Anazy RM, Javaid K, Al-Mandeel HM. Linguistic and clinical validation of the Arabic-translated Aberdeen Menorrhagia Severity Scale as an indicator

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

of quality of life for women with abnormal uterine bleeding. Saudi Med J. 2012;33: 869–74. Available: http://www.ncbi.nlm.nih.gov/pubmed/22886120

184. Hudgens S, Gauthier M, Hunsche E, Kang J, Li Y, Scippa K, et al. Development of the Bleeding and Pelvic Discomfort Scale for Use in Women With Heavy Menstrual Bleeding Associated With Uterine Fibroids. Value in Health. 2022;25. doi:10.1016/j.jval.2022.06.005

185. Nguyen AM, Humphrey L, Kitchen H, Rehman T, Norquist JM. A qualitative study to develop a patient-reported outcome for dysmenorrhea. Quality of Life Research. 2015;24: 181–191. doi:10.1007/s11136-014-0755-z

186. Guan Y, Nguyen AM, Wratten S, Randhawa S, Weaver J, Arbelaez F, et al. The endometriosis daily diary: qualitative research to explore the patient experience of endometriosis and inform the development of a patient-reported outcome (PRO) for endometriosis-related pain. J Patient Rep Outcomes. 2022;6: 5. doi:10.1186/s41687-021-00409-8

187. Wyrwich KW, O'Brien CF, Soliman AM, Chwalisz K. Development and Validation of the Endometriosis Daily Pain Impact Diary Items to Assess Dysmenorrhea and Nonmenstrual Pelvic Pain. Reproductive Sciences. 2018;25: 1567–1576. doi:10.1177/1933719118789509

188. Moradi M, Parker M, Sneddon A, Lopez V, Ellwood D. The Endometriosis Impact Questionnaire (EIQ): a tool to measure the long-term impact of endometriosis on different aspects of women's lives. BMC Womens Health. 2019;19: 64. doi:10.1186/s12905-019-0762-x

189. van Nooten FE, Cline J, Elash CA, Paty J, Reaney M. Development and content validation of a patient-reported endometriosis pain daily diary. Health Qual Life Outcomes. 2018;16: 3. doi:10.1186/s12955-017-0819-1

190. Namazi M, Zareiyan A, Jafarabadi M, Behboodi Moghadam Z. Endometriosis reproductive health questionnaire (ERHQ): A self-administered questionnaire to measure the reproductive health in women with endometriosis. J Gynecol Obstet Hum Reprod. 2021;50: 101860. doi:10.1016/j.jogoh.2020.101860

191. Ahmadpour P, Jahangiry L, Bani S, Iravani M, Mirghafourvand M. Validation of the Iranian version of the ENDOPAIN-4D questionnaire for measurement of painful symptoms of endometriosis. J Obstet Gynaecol (Lahore). 2022;42. doi:10.1080/01443615.2022.2049726

192. Fauconnier A, Staraci S, Daraï E, Descamps P, Nisolle M, Panel P, et al. A self-administered questionnaire to measure the painful symptoms of endometriosis: Results of a modified DELPHI survey of patients and physicians. J Gynecol Obstet Hum Reprod. 2018;47: 69–79. doi:10.1016/j.jogoh.2017.11.003

193. Puchar A, Panel P, Oppenheimer A, Du Cheyron J, Fritel X, Fauconnier A. The ENDOPAIN 4D Questionnaire: A New Validated Tool for Assessing Pain in Endometriosis. J Clin Med. 2021;10: 3216. doi:10.3390/jcm10153216

194. As-Sanie S, Laufer MR, Missmer SA, Murji A, Vincent K, Eichner S, et al. Development of a visual, patient-reported tool for assessing the multi-dimensional burden of endometriosis. Curr Med Res Opin. 2021;37: 1443–1449. doi:10.1080/03007995.2021.1929896

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

195. Deal LS, Williams VSL, Fehnel SE. Development of an Electronic Daily Uterine Fibroid Symptom Diary. The Patient: Patient-Centered Outcomes Research. 2011;4: 31–44. doi:10.2165/11537290-000000000-00000

196. Olliges E, Bobinger A, Weber A, Hoffmann V, Schmitz T, Popovici RM, et al. The Physical, Psychological, and Social Day-to-Day Experience of Women Living With Endometriosis Compared to Healthy Age-Matched Controls—A Mixed-Methods Study. Front Glob Womens Health. 2021;2: 767114. doi:10.3389/fgwh.2021.767114

197. Bushnell DM, Martin ML, Moore KA, Richter HE, Rubin A, Patrick DL. Menorrhagia Impact Questionnaire: assessing the influence of heavy menstrual bleeding on quality of life. Curr Med Res Opin. 2010;26: 2745–55. doi:10.1185/03007995.2010.532200

198. Pike M, Chopek A, NL Y, Usuba K, MJ B, McLaughlin R, et al. Quality of life in adolescents with heavy menstrual bleeding: Validation of the Adolescent Menstrual Bleeding Questionnaire (aMBQ). Res Pract Thromb Haemost. 2021;5: e12615. doi:10.1002/rth2.12615

199. Rezende GP, Brito LGO, Gomes DAY, Souza LM de, Polo S, Benetti-Pinto CL. Assessing a cut-off point for the diagnosis of abnormal uterine bleeding using the Menstrual Bleeding Questionnaire (MBQ): a validation and cultural translation study with Brazilian women. Sao Paulo Med J. 2023;142: e2022539. doi:10.1590/1516-3180.2022.0539.R2.100423

200. Rodpetch T, Manonai J, Angchaisuksiri P, Boonyawat K. A quality-of-life questionnaire for heavy menstrual bleeding in Thai women receiving oral antithrombotics: Assessment of the translated Menstrual Bleeding Questionnaire. Res Pract Thromb Haemost. 2021;5: e12617. doi:10.1002/rth2.12617

201. Lee AM, So-Kum Tang C, Chong C. A culturally sensitive study of premenstrual and menstrual symptoms among Chinese women. Journal of Psychosomatic Obstetrics & Gynecology. 2009;30: 105–114. doi:10.1080/01674820902789241

202. Moos RH. The development of a menstrual distress questionnaire. Psychosom Med. 1968;30: 853–67. doi:10.1097/00006842-196811000-00006

203. Cassioli E, Rossi E, Melani G, Faldi M, Rellini AH, Wyatt RB, et al. The menstrual distress questionnaire (MEDI-Q): reliability and validity of the English version. Gynecological Endocrinology. 2023;39. doi:10.1080/09513590.2023.2227275

204. Vannuccini S, Rossi E, Cassioli E, Cirone D, Castellini G, Ricca V, et al. Menstrual Distress Questionnaire (MEDI-Q): a new tool to assess menstruation-related distress. Reprod Biomed Online. 2021;43: 1107–1116. doi:10.1016/j.rbmo.2021.08.029

205. Parker MA, Kent AL, Sneddon A, Wang J, Shadbolt B. The Menstrual Disorder of Teenagers (MDOT) Study No. 2: Period ImPact and Pain Assessment (PIPPA) Tool Validation in a Large Population-Based Cross-Sectional Study of Australian Teenagers. J Pediatr Adolesc Gynecol. 2022;35: 30–38. doi:10.1016/j.jpag.2021.06.003

206. Calaf J, Cancelo MJ, Andeyro M, Jiménez JM, Perelló J, Correa M, et al. Development and Psychometric Validation of a Screening Questionnaire to Detect Excessive Menstrual Blood Loss That Interferes in Quality of Life: The SAMANTA Questionnaire. J Womens Health (Larchmt). 2020;29: 1021–1031. doi:10.1089/jwh.2018.7446

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

207. Perelló-Capo J, Rius-Tarruella J, Andeyro-García M, Calaf-Alsina J. Sensitivity to Change of the SAMANTA Questionnaire, a Heavy Menstrual Bleeding Diagnostic Tool, after 1 Year of Hormonal Treatment. J Womens Health. 2023;32. doi:10.1089/jwh.2022.0155

208. Pérez-Campos E, Dueñas JL, de la Viuda E, Gómez MÁ, Lertxundi R, Sánchez-Borrego R, et al. Development and validation of the SEC-QOL questionnaire in women using contraceptive methods. Value Health. 2011;14: 892–9. doi:10.1016/j.jval.2011.08.1729

209. Perelló J, Pujol P, Pérez M, Artés M, Calaf J. Heavy Menstrual Bleeding-Visual Analog Scale, an Easy-to-Use Tool for Excessive Menstrual Blood Loss That Interferes with Quality-of-Life Screening in Clinical Practice. Womens Health Rep (New Rochelle). 2022;3: 483–490. doi:10.1089/whr.2021.0139

210. Teherán A, Pineros LG, Pulido F, Mejía Guatibonza MC. WaLIDD score, a new tool to diagnose dysmenorrhea and predict medical leave in university students. Int J Womens Health. 2018;10: 35–45. doi:10.2147/IJWH.S143510

211. Dimentberg E, Cardaillac C, Richard E, Plante A-S, Maheux-Lacroix S. Translation and Cultural Validation of the WERF EPHect Endometriosis Patient Questionnaire into Canadian French. Journal of Obstetrics and Gynaecology Canada. 2021;43: 817–821. doi:10.1016/j.jogc.2021.03.019

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

1 ADDITIONAL TABLE

2 Table 4: List of full instruments and characteristics

Full Name of instrument	Available Languages	Available Electronically?*	Who fills out instrument?	BLEEDING				BLOOD			PAIN	PERCEP-TIONS	Quality Score	Utility Score	Evidence Score**	References
				Duration	Volume	Frequency	Regularity	Color	Consistency	Smell						
Instruments that Measure Bleeding and/or Blood (n=13)																
Alkaline Hematin Assay	NA	No	Patient/Participant		X								2.00	2.00	8†	[108]
Daily Diary, Menstrual Cycle Length	English	Yes	Patient/Participant	X	X								1.25	2.00	14†	[70]
Daily Diary, Menopause Classification‡	English, Cantonese, Japanese, Spanish	No	Patient/Participant	X		X							2.00	2.50‡	16	[69]
Mansfield-Voda-Jorgensen Menstrual Bleeding Scale	English	No	Patient/Participant		X								2.00	3.00	9	[109]
Menstrual Blood Loss Score Questionnaire	Spanish	No	Patient/Participant	X	X								2.25	2.67	17	[71]
Menstrual Collection	English, Icelandic	No	Patient/Participant		X								2.67	1.53	9	[110–114]
Menstrual Record and Recall	English	No	Patient/Participant		X								2.00	2.67	14	[115]
Pictorial Blood Loss Assessment Charts & Menstrual Pictograms†† ‡	Dutch, English, German, Norwegian	No	Patient/Participant		X				X				2.67††	2.00‡	84	[116–126]
Prospective Self Report, Menstrual Regularity	Not Reported	No	Patient/Participant	X									2.00	2.33	13†	[68]
Quantitative Model for Menstrual Blood Loss††	Multi-Site Study	No	Researcher		X								2.83††	1.67	16	[67]
Retrospective Self-Report, Last Menstrual Period	English	No	Patient/Participant			X							1.67	3.00	14	[78]
Retrospective Self Report, Menstrual Length (Small & Jukic)	English	No	Patient/Participant			X							2.33	2.83	30	[76,77]
Retrospective Self Report, Menstrual Length (Bachand)	English	No	Patient/Participant			X							2.00	3.00	13	[79]
Instruments that Measure Pain (n=4)																
Numeric Rating Scale	English, Portuguese, Spanish	Yes	Patient/Participant								X		2.60	2.00	30	[127,128]
Pain Drawing	Portuguese	Yes	Patient/Participant								X		2.67	3.00	17	[87]
Retrospective Self Report, Menstrual Discomfort	English	No	Patient/Participant								X		2.50	3.00	14	[88]
Squeezing Pain Bulb	English	Yes	Patient/Participant								X		1.50	1.00	8	[86]
Visual Analogue Scales: Pain	Multi-Site Study	No	Patient/Participant								X		2.00	2.00	20	[129,130]
Instruments that Measure Perceptions (n=19)																
Adolescent Dysmenorrhic Self-Care Scale ‡	Cantonese, Mandarin	No	Patient/Participant									X	3.00	2.88‡	45	[131,132]
Dysmenorrhea Symptom Interference Scale	English	Yes	Patient/Participant									X	2.80	3.00	20	[133]
Endometriosis Health Profile-30††	Chinese, Danish, Dutch, English, French, Italian, Portuguese, Portuguese (Brazilian), Malay, Norwegian, Swedish, Turkish, Persian	Yes (French)	Patient/Participant									X	3.00††	2.52	332	[65,66,134–151]
Endometriosis Health Profile-5	Croatian, English, French	Yes (Croatian)	Patient/Participant									X	3.00	3.00	53	[152–155]

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC BY-NC-ND 4.0 International license .

Full Name of instrument	Available languages	Electronically?	Who fills out instrument?	BLEEDING				BLOOD			PAIN	PERCEP-TIONS	Quality Score	Utility Score	Evidence Score**	References
				Duration	Volume	Frequency	Regularity	Color	Consistency	Smell						
Endometriosis Impact Scale‡	English, French, German	Yes	Patient/Participant									X	3.00	2.75‡	17	[156]
Endometriosis Treatment Satisfaction Questionnaire	English	No	Patient/Participant									X	2.75	3.00	20	[157]
Functional and Emotional Measure of Dysmenorrhea	Chinese	No	Patient/Participant									X	2.25	0.00	9	[158]
Injustice Experience Questionnaire-Chronic and the Contribution of Perceived Injustice	Japanese	Yes	Patient/Participant									X	2.00	2.67	14	[159]
(Menorrhagia) Multi-Attribute Utility Score	English	No	Patient/Participant									X	2.40	2.50	25	[160–162]
Menstrual Attitudes Questionnaire	Bengali, English, Greek, Nepali, Turkish	No	Patient/Participant									X	2.25	1.73	29	[163–167]
Menstrual Health Seeking Behaviors Questionnaire	Persian	No	Patient/Participant									X	2.75	2.00	15	[168]
Menstrual Hygiene Management Scale	Hindi	Yes	Patient/Participant									X	2.33	3.00	10	[93]
Menstrual Joy Questionnaire	English	No	Patient/Participant									X	1.50	3.00	9	[169]
Menstrual Practices Questionnaire	English	No	Patient/Participant									X	2.60	2.50	18	[170]
Menstrual Self-Evaluation Scale	English	No	Patient/Participant									X	2.00	2.50	11	[171]
Menstruation-Related, Activity Restriction Questionnaire	English, Hindi	No	Patient/Participant									X	2.00	2.67	14	[92]
Military Women's Attitudes Toward Menstrual Suppression Scale	English	No	Patient/Participant									X	2.50	1.67	15	[172]
Uterine Fibroid Symptom and Quality of Life Questionnaire‡	Chinese, Dutch, English, Portuguese, Spanish	Yes (Dutch)	Patient/Participant; Researcher									X	2.80	2.67‡	164	[72,173–180]
Working Stressors and Coping Strategies Associated with Menstrual Symptoms among Nurses	Not Reported	Yes	Patient/Participant									X	2.75	1.67	16	[181]
World Health Organization Disability Assessment Schedule 2.0	Portuguese	Yes	Patient/Participant									X	2.00	0.00	4	[182]
Instruments that Measure Multiple CIMCs (n=28)																
Aberdeen Menorrhagia Severity Scale‡	Arabic, English	No	Patient/Participant	X	X	X	X		X		X	X	2.50	2.88‡	19	[80,183]
Bleeding and Pelvic Discomfort Scale	English	No	Patient/Participant								X	X	2.80	3.00	23	[184]
electronic Personal Assessment Questionnaire - Menstrual, Pain, and Hormonal	English	Yes	Patient/Participant	X	X		X		X		X	X	2.00	2.00	14	[83]
Dysmenorrhea Daily Diary††	English	Yes	Patient/Participant	X	X						X	X	2.67††	2.17	34	[75,185]
Endometriosis Daily Diary	English, Cantonese, Japanese, Spanish	Yes	Patient/Participant								X	X	1.83	2.00	17	[186]
Endometriosis Daily Pain Impact Diary	English	Yes	Patient/Participant								X	X	2.80	2.33	21	[187]
Endometriosis Impact Questionnaire‡	English	Yes	Patient/Participant		X		X				X	X	2.75	2.50‡	21	[188]
Endometriosis Pain and Bleeding Diary	English	Yes	Patient/Participant		X						X	X	2.75	2.00	17	[89]
Endometriosis Pain Daily Diary	English, Japanese	Yes	Patient/Participant								X	X	3.00	2.33	13	[189]

medRxiv preprint doi: https://doi.org/10.1101/2024.04.04.24305348; this version posted April 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC BY-NC-ND 4.0 International license .

Full Name of instrument	Available languages	Electronically?	Who fills out instrument?	BLEEDING				BLOOD			PAIN	PERCEP-TIONS	Quality Score	Utility Score	Evidence Score**	References
				Duration	Volume	Frequency	Regularity	Color	Consistency	Smell						
Endometriosis Reproductive Health Questionnaire	Persian	No	Patient/Participant		X						X	X	2.25	2.50	14	[190]
Endometriosis Self-Assessment Tool	Korean	No	Patient/Participant		X				X		X	X	3.00	2.67	20	[84]
Endometriosis Symptom Diary‡	English, French, German	Yes	Patient/Participant		X						X	X	3.00	2.25‡	15	[156]
ENDOPAIN-4D‡	French, Persian	No	Patient/Participant								X	X	2.80	2.29‡	52	[191–193]
Endowheel‡	English	No	Patient/Participant		X		X				X	X	3.00	2.50‡	16	[194]
Fibroid Symptom Diary	English	Yes	Patient/Participant		X				X		X	X	2.50	2.00	11	[195]
Measure Compilation (Olliges)	German	Yes	Patient/Participant		X						X	X	2.00	1.67	11	[196]
Menorrhagia Impact Questionnaire	English	No	Patient/Participant		X							X	3.00	2.50	20	[197]
Menstrual Bleeding Questionnaire††	English, Portuguese, Thai	No	Patient/Participant	X	X		X		X		X	X	3.00††	2.33	60	[73,198–200]
Menstrual Distress Questionnaire (Moos)	English	No	Patient/Participant								X	X	2.25	2.33	16	[201,202]
Menstrual Distress Questionnaire (Vannuccini)	English, Italian	Yes	Patient/Participant								X	X	2.75	2.50	33	[203,204]
Menstrual Health Instrument	Korean	No	Patient/Participant			X	X				X	X	2.60	2.50	18	[82]
Menstrual Insecurity Tool	Oriya (Odia)	No	Patient/Participant				X			X	X	X	2.75	3.00	17	[85]
New Zealand Survey of Adolescent Girls' Menstruation	English	Yes	Patient/Participant	X	X	X	X		X		X	X	2.33	1.33	11	[90]
Period ImPact and Pain Assessment	English	No	Patient/Participant								X	X	2.00	3.00	12	[205]
PERIOD-QOL	English	Yes	Patient/Participant	X	X						X	X	2.75	2.00	15	[74]
SAMANTA Questionnaire	Spanish	No	Patient/Participant	X	X							X	3.00	3.00	35	[206,207]
Spanish Society of Contraception Quality-of-Life	Spanish	No	Patient/Participant								X	X	3.00	3.00	24	[208]
Visual Analogue Scales: Bleeding	Spanish	No	Patient/Participant		X							X	2.00	2.50	9	[209]
Working Ability, Location, Intensity, Days of Pain, Dysmenorrhea Score	Spanish	No	Patient/Participant								X	X	1.75	2.00	13	[210]
World Endometriosis Research Foundation Endometriosis Phenome and Biobanking Harmonisation Project Standard Questionnaire‡	English, French	No	Patient/Participant	X	X	X	X				X	X	2.25	1.50‡	21	[81,211]
Total Number of full Instruments	**68**		**Total/Average**	**12**	**26**	**8**	**9**	**0**	**7**	**1**	**32**	**50**	**2.44**	**2.33**	**25.59**	

* According to publications, "Yes" indicates either fully or partly electronic
** Sum of quality and utility scores for studies conducted after 2006.
† Evidence score based on only one study before 2006
†† Score provided for every aspect of quality (no scores of 0 in any category)
‡ Score provided for every aspect of utility (no scores of 0 in any category)