





Development of measures for assessing mistreatment of women during facility-based childbirth based on labour observations

Blair O Berger ¹, Donna M Strobino,¹ Hedieh Mehrdash,² Meghan A Bohren ³, Kwame Adu-Bonsaffoh,⁴ Hannah H Leslie ⁵, Theresa Azonima Irinyenikan,⁶ Thae Maung Maung ⁷, Mamadou Dioulde Balde,⁸ Özge Tunçalp⁹

To cite: Berger BO, Strobino DM, Mehrdash H, *et al.* Development of measures for assessing mistreatment of women during facility-based childbirth based on labour observations. *BMJ Global Health* 2021;**5**:e004080. doi:10.1136/bmjgh-2020-004080

Handling editor Seye Abimbola

► Additional material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjgh-2020-004080>).

Received 29 September 2020
Revised 3 March 2021
Accepted 4 March 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to
Dr Blair O Berger;
blair.berger@jhu.edu

ABSTRACT

Introduction Mistreatment of women during childbirth is increasingly recognised as a significant issue globally. Research and programmatic efforts targeting this phenomenon have been limited by a lack of validated measurement tools. This study aimed to develop a set of concise, valid and reliable multidimensional measures for mistreatment using labour observations applicable across multiple settings.

Methods Data from continuous labour observations of 1974 women in Nigeria (n=407), Ghana (n=912) and Guinea (n=655) were used from the cross-sectional WHO's multicountry study 'How women are treated during facility-based childbirth' (2016–2018). Exploratory factor analysis was conducted to develop a scale measuring interpersonal abuse. Two indexes were developed through a modified Organisation for Economic Co-operation and Development approach for generating composite indexes. Measures were evaluated for performance, validity and internal reliability.

Results Three mistreatment measures were developed: a 7-item Interpersonal Abuse Scale, a 3-item Exams & Procedures Index and a 12-item Unsupportive Birth Environment Index. Factor analysis results showed a consistent unidimensional factor structure for the Interpersonal Abuse Scale in all three countries based on factor loadings and interitem correlations, indicating good structural construct validity. The scale had a reliability coefficient of 0.71 in Nigeria and approached 0.60 in Ghana and Guinea. Low correlations (Spearman correlation range: -0.06–0.19; p>0.05) between mistreatment measures supported our decision to develop three separate measures. Predictive criterion validation yielded mixed results across countries. Both items within measures and measure scores were internally consistent across countries; each item co-occurred with other items in a measure, and scores consistently distinguished between 'high' and 'low' mistreatment levels.

Conclusion The set of concise, comprehensive multidimensional measures of mistreatment can be used in future research and quality improvement initiatives targeting mistreatment to quantify burden, identify risk factors and determine its impact on health and well-being outcomes. Further validation and reliability testing of the measures in other contexts is needed.

Key questions

What is already known?

- There is growing recognition that mistreatment of women during facility-based childbirth is a widespread issue impacting women's experiences of care, and it is a deterrent to future care seeking.
- There remains a lack of consensus on operational definitions of measures and constructs of mistreatment, as well as validated measures.

What are the new findings?

- Using data from continuous labour observations of women in Nigeria, Ghana and Guinea from the WHO multicountry study 'How women are treated during facility-based childbirth' we developed a set of three concise mistreatment measures: a 7-item Interpersonal Abuse Scale, a 3-item Exams & Procedures Index and a 12-item Unsupportive Birth Environment Index.
- The three mistreatment measures showed strong preliminary construct validity and internal consistency.
- Further validation and reliability testing of the measures in other contexts is needed.

What do the new findings imply?

- Validated measures are essential to quantify the scale of the mistreatment as a global issue, identify risk factors and determine its impact on health outcomes.
- Our study presents a set of three short measures of mistreatment that can be used as internally consistent measures with adequate preliminary validity to assess multiple dimensions of mistreatment during childbirth.
- These measures can be adapted and used in future research and quality improvement initiatives to quantify the burden, frequency and overlap of multiple types of mistreatment in a standardised way that can be compared across studies, settings and time periods.

INTRODUCTION

There have been significant declines globally in maternal morbidity and mortality following strategies that have targeted resource

availability, logistics and access, and strengthening maternity care systems on the premise that if women reach facilities, many barriers to improving maternal health outcomes would be overcome.¹⁻³ However, previous research has underscored that high coverage of essential technical interventions alone does not yield expected improvements in maternal health; rather, a focus on measuring and improving quality of care, including interpersonal aspects and experiences of care, are critical.⁴⁻⁶ A recent *Lancet Global Health* Commission on High-Quality Health Systems concluded that poor-quality care has become a larger barrier to reducing mortality than inadequate access, and most health system measures fail to capture metrics that are important to people, including user experience, confidence in the health system and health outcomes.⁷⁻⁹

There is growing recognition globally that a key determinant for women seeking facility-based obstetric care is how they are treated during childbirth. Several terms have been used to describe poor care and abusive, disrespectful, negligent or discriminatory treatment of women giving birth in facilities (hereafter referred to in short as ‘mistreatment’).¹⁰⁻¹² These terms have been framed as a subset of the larger issues of violence against women, human rights violations, quality of care, health systems issues or a combination of these.¹³ Bowser and Hill’s landmark landscape analysis was the first to review existing evidence and convene an expert working group to develop a classification system for ‘disrespect and abuse during facility-based childbirth’.¹⁴ A global mixed-methods systematic review of mistreatment by Bohren *et al.* (2015) cited widespread mistreatment in 34 countries and resulted in the first evidence-based typology of mistreatment (the WHO typology), which included domains of physical abuse, verbal abuse, sexual abuse, stigma and discrimination, failure to meet professional standards of care, poor rapport and communication between women and providers, and health systems conditions and constraints.¹²

Recent prevalence estimates of mistreatment range widely across studies from 13% to 98%.^{11 15-29} Definitions of mistreatment, measurement tools, study designs, data collection modes and study samples vary significantly, posing challenges in comparing both the findings and the validity of the measurement approaches. Mistreatment is often measured as a binary outcome of experiencing at least one kind of mistreatment, particularly when risk factors are assessed.^{17 19 30-35} This single binary indicator captures a wide range of behaviours and events included in instruments to characterise mistreatment, thus limiting the ability to distinguish different types of mistreatment and to assess overlap and co-occurrence of these experiences.³⁶⁻⁴⁰ There has also been growing momentum to develop tools that capture women’s experiences during childbirth. Two scales were recently developed to measure perceptions of respectful maternity care and person-centred care.^{22 41-43} However, the focus of the scales is broader than specifically capturing forms of

mistreatment as they seek to situate respectful maternity care as a core feature of person-centred care.^{22 41-43}

There is also debate over a standard mode of assessment for mistreatment, with previous studies using direct labour observations, facility exit interviews or women’s reports via community surveys.^{11 16 28 44-46} Both labour observations and women’s reports offer important perspectives that contribute to understanding the forms and magnitude of mistreatment from unique vantage points. Two recent studies assessing discordance between self-report and labour observation reports of mistreatment in Kenya and India found that observers reported significantly higher mistreatment than women, with the authors acknowledging a complex array of social norms, expectations of care, power dynamics and recall issues during birth that may contribute to women under-reporting mistreatment.^{44 46}

Quantitative studies have been critical in providing preliminary estimates of the burden of mistreatment across contexts. However, the evidence base underpinning them has been hampered by lack of consensus on operational definitions and constructs of mistreatment, as well as validated measures. Validated measures are essential to quantify the scale of the problem, identify risk factors and determine its impact on health outcomes. This analysis aimed to develop a set of concise, valid and reliable measures for mistreatment of women during facility-based childbirth in the following dimensions: interpersonal abuse, inappropriate conduct of exams and procedures, and unsupportive birth environment, using labour observation data of women in three countries (Nigeria, Ghana and Guinea).

METHODS

Patient and public involvement

A technical consultation with representatives from advocacy groups, non-governmental organisations, research organisations, universities, professional associations and United Nations agencies was held at the WHO in November 2013, and informed the research questions and design of survey instruments in the WHO study from which data for this analysis were obtained. Women who recently gave birth were involved in exploratory formative research informing tool development, content validity testing and providing feedback on the validity testing of the community survey tool.⁴⁷⁻⁵⁴

Data source and study participants

This secondary data analysis used data from the cross-sectional WHO multicountry study ‘How women are treated during facility-based childbirth,’ conducted in two phases between 2014 and 2018 in four purposively selected countries, Ghana, Guinea, Nigeria and Myanmar. The first study phase was involved a mixed-methods systematic review and primary qualitative research to develop two measurement instruments for mistreatment, a direct labour observation tool and a postpartum community

survey tool. The second study phase implemented and validated the instruments. Details of the WHO study methodology are described elsewhere.^{10 47 48} Data for this analysis were obtained from the labour observation tool from phase 2 of the study conducted in Nigeria, Ghana and Guinea (labour observations were not conducted in Myanmar).

A total of 2016 women were observed during childbirth in three purposively-selected public, urban, secondary-level or higher facilities in each country. Observations occurred in Nigeria from September 2016 to February 2017 and in Ghana and Guinea from August 2017 to February 2018. This analysis included 1974 women with complete data on the mistreatment items (407 women in Nigeria, 912 in Ghana and 655 women in Guinea; 42 women were excluded due to missing data (n=1 in Nigeria, n=14 in Ghana and n=27 in Guinea)). All pregnant women entering study facilities without visible signs of distress or obstetric emergency were approached and screened for eligibility on admission. Women were eligible if they were in active labour and admitted to the facility for childbirth, were at least 15 years of age and provided written consent. Participating women were continuously observed by a trained, non-clinical female observer from admission to 2 hours postpartum. Observers completed a structured instrument collecting information on maternal sociodemographic characteristics and reproductive history, use of medical interventions, maternal and neonatal outcomes, and all incidents of mistreatment between women, providers and facility staff. Observers recorded four aspects of a mistreatment incident: form of mistreatment, the time it occurred, whether it occurred during the intrapartum or postpartum period, and who committed it; a separate incident report was recorded for each incident in the event of repeated occurrences of the same type of mistreatment.⁴⁷

Development and preliminary validation of measurement tools

The WHO typology of mistreatment served as the basis for items included in the WHO survey instruments and the item pool for the mistreatment measures developed in this analysis.¹² Final item development, including face and content validation of the study instruments occurred through expert review by a group of global maternal health experts to determine the relevance of each item to the desired construct and suggestions for additional items where needed. The labour observation instrument was then piloted with a convenience sample of women in a single study site in Nigeria.⁴⁷

Dimensions of mistreatment

We use the term ‘mistreatment’ for the incidents represented in the measures, as this terminology underscores that mistreatment can be intentional or unintentional, and occurs at two levels: (1) the interpersonal level between provider and patient and (2) the systems level through failures of the health facility and health system.¹²

We posit that the WHO typology contains domains of mistreatment that do not share a single common underlying factor. We thus operationalised mistreatment as capturing a spectrum of behaviours and experiences. The latent construct underlying physical abuse, verbal abuse and stigma and discrimination is related to interpersonal abuse within a more general violence framework, while the latent construct underlying failures to meet professional standards of care, poor rapport between women and providers, and health systems conditions and constraints is intrinsically tied to mistreatment in the process of care within a broader quality of care framework.

Statistical methods

All analyses were conducted using Stata V.15.0.⁵⁵ More detailed information on the analytical approach is provided in online supplemental file.

Item construction and identifying dimensions

A preliminary pool of 56 binary items in the labour observation tool was used to develop the mistreatment measures, based on five domains of the WHO typology. Sexual abuse was not captured in the WHO study instruments. We did not include stigma and discrimination because preliminary data evaluation indicated a uniformly low frequency of items in this domain across the three country samples.

Extensive exploratory data analysis on the three country samples involved an inductive approach to assess item distributions and to determine how to best operationalise dimensions of mistreatment in the measures. We iteratively assessed potential combinations of items based on conceptual consistency, practicality and understandability across study countries in an effort to maintain face and content validity. All mistreatment items were constructed as binary (‘0=no mistreatment’ and ‘1=mistreatment’). Where applicable, items were reverse coded to be consistent with this structure. Responses on ‘not applicable’ or ‘don’t know’ categories were coded as 0, rather than omitting this data, to give the most conservative estimate of the mistreatment measures.

We developed and assessed the validity and reliability of a new set of concise measures covering three dimensions of mistreatment, identified on both empirical and conceptual bases: interpersonal abuse, inappropriate conduct of exams and procedures (referred to hereafter as ‘exams & procedures’) and unsupportive birth environment. While we initially planned to develop a single psychometric scale or a set of subscales for mistreatment, based on a series of tests (Bartlett’s test of sphericity and Kaiser-Meyer-Olin test for sampling adequacy), factorability was only established for items in the interpersonal abuse dimension. The final set of scored measures for mistreatment thus included one psychometric scale measuring interpersonal abuse (a more reflexive indicator where items are theoretically homogenous and highly correlated), and two indexes measuring exams

& procedures and unsupportive birth environment (formative indicators where items have different underlying causes, no assumption of homogeneity, and each item contributes uniquely to the level of a measured construct).^{56 57}

Interpersonal Abuse Scale development: psychometric analysis

The Interpersonal Abuse Scale was developed using a psychometric approach outlined by DeVellis and Netermeyer *et al.*^{58 59} Exploratory factor analysis (EFA) was conducted separately by country to assess the construct validity and reliability of the scale.^{58 59} Given the early developmental stages of operationalising the Interpersonal Abuse Scale as a dimension of mistreatment, confirmatory factor analysis (CFA) was yet not warranted because CFA tests a strong theoretical a priori conceptualisation and factor structure of a construct.^{37 38 58 59} A principal components analysis (PCA) was conducted on tetrachoric correlation matrices to determine the number of common factors to extract for factor analysis based on four criteria: Kaiser's rule of retaining eigenvalues > 1.0, the 'bend' in the scree plot (plot of eigenvalues), proportion of variance explained by the factors and parallel analysis tests.^{58–60} EFA was then performed using iterated principal factor estimation. To establish unidimensionality, items with either low ($\lambda < 0.40$) or very high ($\lambda > 0.90$, indicating possible redundancy) factor loadings were omitted.^{58 59}

Two types of construct validity during psychometric analyses were assessed in three ways: (1) standardised loadings of > 0.40 on a single dominant factor, coupled with, (2) high interitem correlations provided evidence of structural construct validity and (3) a consistent factor structure across multiple country samples further signalled structural and cross-cultural construct validity/measurement invariance.^{60–62} Corrected item-to-total correlations, interitem correlations and the Kuder-Richardson-20 (KR-20) coefficient (the corollary to Cronbach's alpha used for scales based on dichotomous items) were used to assess scale reliability.^{37 38 58 59} A range of 0.15–0.50 corrected item-to-total and interitem correlations was used as evidence of adequate internal consistency, which has been posited as an acceptable range for new measures in early stages of scale development.^{58–60}

An approach informed by clinimetrics was used to develop the Exams & Procedures Index and an Unsupportive Birth Environment Index in the three settings using a method adapted from the Organisation for Economic Co-operation and Development (OECD) procedure for composite index development,⁶³ the indexing approach outlined in the Consensus-based Standards for selection of health Measurement Instruments initiative (COSMIN),^{61 62} and existing indices measuring quality of care in maternal and reproductive health.^{64–68} The Exams & Procedures Index was constructed using the denominator of women who received at least one vaginal examination and/or at least one procedure

(caesarean, episiotomy, hysterectomy, tubal ligation or postpartum intrauterine device (IUD) insertion). Items occurring in > 90% of observations in all three countries were omitted because they would limit the ability for the measures to discriminate between levels of mistreatment, and they would artificially inflate index scores. Infrequently occurring items (< 5%) in all three settings were considered for deletion but were ultimately retained to improve content validity and err on the side of overinclusion based on expert feedback and their limited effect on scoring; a primary goal in indexing is to include enough items to fully cover the content of the measured dimension since formative items define the latent construct.^{69 70}

Items in all three mistreatment measures were scored as '1' if there was a least one report of the incident, and '0', otherwise, where higher scores indicate higher presence of different types of mistreatment. Scores were aggregated separately by measure using simple summative scoring without weighting. The range of scores by measure was: Interpersonal Abuse Scale (0–7), Exams & Procedures Index (0–3) and Unsupportive Birth Environment Index (0–12).

Measure validation and reliability assessment

Construct validation of the three measures was assessed via Spearman-rank correlations between mistreatment measures, using the benchmark of correlations < 0.30 ($p \geq 0.05$) as evidence of significant correlation.⁵⁹ The ability to establish concurrent criterion validity was limited by the lack of a 'gold-standard' measure of mistreatment, as is commonly the case with social and psychological constructs.⁵⁸ Predictive criterion validity was evaluated through bivariate logistic regressions of a global measure of satisfaction with care and intentions to give birth in the same facility in the future (based on linked data of the same women's responses on the WHO community survey tool administered up to 8 weeks postpartum) on ordinal mistreatment measure scores (see online supplemental appendix 1 for details).

Internal consistency of items within each measure was assessed through the percent distributions of co-occurring mistreatment items; the proportion of items that occur with at least one or several other items in a measure tends to be high in internally consistent measures. To determine consistency of measure scores (i.e., whether they can consistently distinguish between 'high' and 'low' scoring groups), the proportions of women experiencing a mistreatment item who had 'high' and 'low' scores were calculated, with levels determined by those scoring above or below the country-specific mean score on the measure.⁷¹

RESULTS

Table 1 shows sociodemographic, provider and obstetric characteristics of the sample by country. The average maternal age at the time of birth was highest in Nigeria (29.3 years); the sample from Guinea (23.8 years) was

Table 1 Individual, provider and birth characteristics of the study sample by country

Country	Nigeria (N=407)	Ghana (N=912)	Guinea (N=655)
	n (%)	n (%)	n (%)
Individual characteristics			
Age (years): mean (SD)	29.3 (5.4)	28.4 (5.9)	23.8 (5.6)
Education level			
None/primary	22 (5.4)	293 (32.1)	438 (66.9)
Secondary	176 (43.2)	462 (50.7)	152 (23.2)
Tertiary	204 (50.1)	134 (14.7)	43 (6.6)
Vocational or other	5 (1.2)	23 (2.5)	22 (3.4)
Marital status			
Married	386 (94.8)	735 (80.6)	612 (93.4)
Not married	21 (5.2)	177 (19.4)	43 (6.6)
Number of previous births			
0	163 (40.1)	343 (37.6)	235 (35.9)
1	110 (27.0)	249 (27.3)	143 (21.8)
2+	134 (32.9)	320 (35.1)	277 (42.3)
Provider characteristics			
Primary labour attendant *			
Nurse/midwife	360 (88.5)	458 (69.9)	762 (83.6)
Doctor	20 (4.9)	103 (15.7)	101 (11.1)
Trainee	22 (5.4)	75 (11.5)	37 (4.1)
None/other	5 (1.2)	19 (2.9)	12 (1.3)
Staff member present at birth*†			
Nurse/midwife	306 (81.4)	667 (74.5)	414 (68.8)
Doctor	66 (17.6)	216 (24.1)	156 (25.9)
Trainee	1 (0.3)	11 (1.2)	32 (5.3)
Other	3 (0.8)	1 (0.1)	--
Characteristics of childbirth			
Mode of birth			
Vaginal	380 (93.4)	765 (83.4)	563 (86.0)
Caesarean	27 (6.6)	147 (16.1)	92 (14.1)
Companion presence during labour and/or childbirth			
Type of companion present at birth (n=163)	n=51	n=82	n=31
Husband/partner	21 (41.2)	62 (75.6)	--
Family/friend	26 (51.0)	20 (24.4)	30 (97.0)
Other	4 (7.8)	--	1 (3.2)
Time of childbirth‡			
Day	201 (49.4)	390 (42.8)	280 (42.8)
Night	206 (50.6)	522 (57.2)	375 (57.3)
Day of childbirth			
Weekday	309 (75.9)	673 (73.8)	513 (78.3)
Weekend	98 (24.1)	239 (26.2)	142 (21.7)

*Trainee=medical students, nursing students, midwifery students.

†Among women with a staff member present at birth.

‡'Day'=8:00–17:00 hours; 'Night'=17:01–7:59 hours.

markedly younger than the samples for Nigeria or Ghana. Women with no or primary education were uncommon in Nigeria (5.4%) where half (50.1%) of women had either postsecondary education or higher, while nearly one-third of women in Ghana and two-thirds of women in Guinea had no or primary education.

Items included in the three measures of mistreatment, the 7-item Interpersonal Abuse Scale, 3-item Exams & Procedures Index and 12-item Unsupportive Birth Environment Index, are noted in [table 2](#). Distributions of mistreatment items, mean and ranges of scores, and the proportion of women scoring above the country-specific mean are shown for the three measures by country in [table 3](#). The Nigerian sample had the highest frequency on all items, and women in this sample scored highest on all three measures, approximately one point or more than women in either Ghana or Guinea. While country-specific means and ranges varied considerably, many women scored higher than their country-specific average on the measures, indicating that multiple, overlapping forms of mistreatment were observed in each dimension.

Psychometric analysis results: Interpersonal Abuse Scale

Seven of the initial 10 items in the physical and verbal abuse domains were retained during factor analysis in all three countries to yield the final Interpersonal Abuse Scale: one item about whether at least one form of physical abuse was observed, and six items about different forms of verbal abuse. While multiple physical abuse items were originally assessed, the low frequency of most of those items required collapsing them into a single item. An item about whether the woman was hissed at was dropped during expert consultations as it was deemed specific to the Nigerian context. Two items were also dropped due to very low frequency and low factor loading ($\lambda < 0.40$): (1) woman was blamed for her or her baby's poor outcome, and (2) woman received any other forms of verbal abuse (not otherwise categorised). The PCA indicated one dominant factor in all three countries based on a single component with an eigenvalue > 1.0 , results of the scree plots and graphical depictions of the parallel analyses (online supplemental appendix 4). [Table 4](#) shows the results of EFA and internal reliability for the Interpersonal Abuse Scale. Factor analysis of the seven items supported a consistent one-factor structure; all items showed strong standardised factor loadings ($\lambda > 0.40$) in all countries, providing evidence of scale unidimensionality and construct validity.

The Interpersonal Abuse Scale showed adequate internal consistency in Nigeria (KR-20 coefficient=0.70). The KR-20 coefficients of 0.57 and 0.54 in Ghana and Guinea, respectively, did not provide evidence for strong internal consistency in these samples. However, with exceptions of 1–2 low pairwise correlations in each country, the inter-item correlations ranged between 0.31–0.62 in Nigeria, 0.29–0.58 in Ghana and 0.23–0.57 in Guinea, indicating internal consistency of this scale in all three countries (online supplemental appendix 2A).

All three items in the Exams & Procedures index were retained. The only item dropped because it occurred in $> 90\%$ of observations in all three countries was an item about whether a woman was asked her preferred birth position (95.1% in Nigeria, 96.5% in Ghana and 97% in Ghana); the study team deemed the conceptual relevance of this item to the construct assessed in the Unsupportive Birth Environment Index was not high enough to retain it. Four items in this index were observed in $< 5\%$ of women in all three countries, but were retained to maintain content validity in this dimension based on team consensus of their relevance to the construct: (1) no interpreter used, (2) whether a woman was neglected, (3) whether a woman was directed to clean up blood or other fluids and (4) whether a woman did not have bed at any time.

Validation analysis

Findings from the factor analysis demonstrated high interitem correlations and a consistent factor structure for a unidimensional Interpersonal Abuse Scale in the three country samples. These results provide evidence for high structural and cross-cultural validity, two important elements of construct validity. Low and non-significant Spearman-rank correlations were noted between each mistreatment measure in all countries (all correlations were < 0.30 , range: -0.06 – 0.19 , $p < 0.05$) (online supplemental appendix 2B), supporting the theoretical approach to developing separate measures to assess the three dimensions of mistreatment (interpersonal abuse, inappropriate conduct of exams and procedures, and unsupportive birth environment); they also provide evidence of construct validity of three measures, rather than a single composite measure. Analysis of predictive criterion validity, however, yielded mixed results for the Interpersonal Abuse Scale and null results for the two indexes across countries (online supplemental appendix 2C).

Reliability: internal consistency analysis

Each item in the measures across all three countries occurred with at least one other item in the measure, indicating internal consistency among the items particularly for the Interpersonal Abuse Scale and the Unsupportive Birth Environment Index (online supplemental appendix 3). Tests of internal consistency showed strong consistency of scores in both the Interpersonal Abuse Scale and the Exams and Procedures Index. The proportion of an observed item was greater among women with 'high' scores than for women with 'low' scores for every item in all measures across settings, indicating the scores consistently distinguished between levels of mistreatment. The Unsupportive Birth Environment Index showed adequate internal consistency of both items and scores, though several items did not vary by level of mistreatment due to low frequency ([table 5](#)).

DISCUSSION

This study used a methodologically rigorous approach to shorten the 56-items WHO labour observation tool⁴⁷

Table 2 Description of mistreatment items

Mistreatment Dimension	Item	WHO mistreatment labour observation tool question	Item scoring coding* (0=no mistreatment 1=mistreatment)
Interpersonal Abuse Scale (7 items)	1.Shouted/screamed at	Was the woman shouted or screamed at?	0: Q=1 (no), 9 (DK) 1: Q=2 (yes)
	2.Insulted	Was the woman insulted?	0: Q=1 (no), 9 (DK) 1: Q=2 (yes)
	3.Scolded	Was the woman scolded?	0: Q=1 (no), 9 (DK) 1: Q=2 (yes)
	4.Mocked	Was the woman mocked?	0: Q=1 (no), 9 (DK) 1: Q=2 (yes)
	5.Negative comments	Did the woman receive negative comments about: (1) her physical appearance (including her weight, genitalia, cleanliness or other aspects of a woman's body), (2) her sexual activity, (3) or her baby's appearance (including his/her sex or other aspects of the baby)?	0: Q=1 (no), 9 (DK) 1: Q=2 (yes) on at least one negative comment
	6.Threatened	Was the woman threatened with: (1) physical violence, (2) a poor outcome for her or her baby, (3) withholding care from her or her baby or (4) with the use of a medical procedure (such as episiotomy, caesarean or other procedure)?	0: Q=1 (no), 9 (DK) 1: Q=2 (yes) on at least one threat
	7.Physical abuse	Did the woman the woman experience at least one of the following (assessed as separate items): Physically struck: pinched, kicked, slapped, punched or hit with instrument, or Forceful restraint or pressure: gagged, physically tied to the bed, held down to the bed forcefully, given forceful downward pressure on abdomen, other use of physical force	0: Q=1 (no), 9 (DK) 1: Q=2 (yes) on at least one physical abuse item
	<i>Deleted Items</i>		<i>Reason for deletion</i>
	<i>Hissed</i>	<i>Was the woman hissed at?</i>	<i>Nigerian-specific item</i>
	<i>Blamed</i>	<i>Was the woman blamed for her or her baby's poor outcome?</i>	<i>Low frequency, low factor loading</i>
Exams & Procedures Index (3 items)	1.Informed consent‡	Did a staff member inform the woman: (1) why a procedure or first vaginal exam performed was needed and (2) obtain her permission for it: 1. caesarean, 2. episiotomy, 3. hysterectomy, 4. tubal ligation, 5. postpartum IUD insertion	0: Q=2 (yes), 9 (DK) 1: Q=1 (no) on either first vaginal exam or procedure
	2.Exposed‡	Did a staff member conduct a vaginal exam in a way that others could see her genitals or breasts?	0: Q=1 (no), 9 (DK) 1: Q=2 (yes)
	3.Confidential information‡	(During a vaginal exam), did a staff member discussed the woman's private health information in a way that others could hear (non-medical staff, other patients or other patients' family members)?	0: Q=1 (no), 9 (DK) 1: Q=2 (yes)
Unsupportive Birth Environment Index (12 items)	<i>Support</i>		
	1.Pain relief	Was the woman offered any form of pain relief or was the woman given pain relief if she requested it?	0: Q=2 (yes), 9 (DK) 1: Q=1 (no)
	2.No interpreter	Was an interpreter used (if the woman's primary birth attendant did not speak same language as the woman)?	0: Q=2 (yes), 9 (DK), 8(NA) 1: Q=1 (no)
	3.No staff present at birth	Was a staff member present when the baby came out?	0: Q=2 (yes), 9 (DK) 1: Q=1 (no)
	4.Not offered labour companion	Was the woman offered to have a labour companion during labour and delivery?	0: Q=2 (yes), 9 (DK) 1: Q=1 (no)
	5.Neglect	Did the woman request medical attention from a health worker that was not responded to?	0: Q=1 (no), 9 (DK) 1: Q=2 (yes)
	<i>Birth environment</i>		
6.Bribe	At any time, did staff suggest or ask the woman (or companion) for a bribe, informal payment or gift?	0: Q=1 (no), 9 (DK) 1: Q=2 (yes)	
7.Clean up blood/fluids	At any time, was the woman instructed to clean up blood, urine, faeces or amniotic fluid?	0: Q=1 (no), 9 (DK). 1: Q=2 (yes)	

Continued

Table 2 Continued

Mistreatment Dimension	Item	WHO mistreatment labour observation tool question	Item scoring coding* (0=no mistreatment 1=mistreatment)
Mistreatment Dimension	Item	WHO Mistreatment Labour Observation Tool Question	Item scoring coding* (Q=0, no mistreatment Q=1, mistreatment)
	8.Fluids	During labour, did the woman have easy access to water or oral fluids?	0: Q=2 (yes), 9 (DK) 1: Q=1 (no)
Unsupportive Birth Environment Index (12 items) (cont.)	9.Mobilise	Was the woman told she could mobilise during labour?	0: Q=2 (yes), 9 (DK) 1: Q=1 (no)
	10.No curtains/partitions	Were curtains, partitions, or other measures used to provide privacy for the woman during: 1. labour, 2. delivery or 3. postpartum?	0: Q=2 (yes), 9 (DK) 1: Q=1 (no) during at least one period
	11.No bed	Did the woman have a bed during: 1. labour, 2. delivery or 3. postpartum?	0: Q=2 (yes), 9 (DK) 1: Q=1 (no) during at least one period
	12.Shared bed	At any time, did the woman have to share a bed with another woman or women?	0: Q=1 (no), 9 (DK) 1: Q=2 (yes)
<i>Deleted Items</i>			<i>Reason for deletion</i>
	<i>Preferred birth position</i>	<i>Was the woman asked her preferred birthing position?</i>	0: Q=2 (yes), 9 (DK) 1: Q=1 (no) >90% frequency in all three countries

Higher scores=higher levels of mistreatment.

*Items constructed from original WHO labour observation tool items (1=No, 2=Yes, 8=Not applicable (NA), 9=Don't know (DK), 'b'=blank).

†Denominator: only those who had a vaginal exam and/or procedure.

‡Denominator: only those who had a vaginal exam.

developed based on five domains of the WHO typology on mistreatment of women during childbirth.¹² A set of three concise, multidimensional measures of mistreatment were developed in this analysis: a 7-item Interpersonal Abuse Scale, a 3-item Exams & Procedures Index and a 12-item Unsupportive Birth Environment Index. The measures showed adequate preliminary internal reliability and construct validity in all three countries. Further validation, particularly of the two indexes, is needed given the mixed criterion validity results.

Three measures were developed rather than a single composite measure, underscoring the complex and multidimensional nature of mistreatment. The validity of this approach was supported by the low correlations between measures across settings. Mistreatment captures a broad range of behaviours and experiences, some of which may reflect more intentional forms of abuse while others reflect poor quality care. The latter may be due to health system deficiencies like inadequate resources, personnel or facility policies; they also may reflect norms in training around pragmatic strategies to establish professional distance or maintain control and compliance during birth in an effort to ensure expedient and/or good birth outcomes.^{13 49 53 72-75} Our measurement approach is consistent with measures operationalised in related research areas. Constructs of violence and abuse are measured using psychometric scales in prior research,⁷⁶⁻⁷⁹ often through adaptations of the Conflict Tactics Scale.^{37 38} Quality of care in maternal and newborn health has typically been measured using clinimetric

indexes and checklists that assess quality of more heterogeneous clinical and service-related items.^{64-66 80}

Separate measures assessing different mistreatment dimensions allow for the ability to distinguish between areas of childbirth care that have higher and lower mistreatment scores to determine tailored quality improvement responses or interventions targeting those high scoring dimensions. For example, high scores on the Interpersonal Abuse Scale or Exams & Procedures Index may be addressed through routine audit and feedback loops or interventions targeting provider behaviour and training, whereas changes to elements of the Unsupportive Birth Index could be made through a systems response focused on resources, infrastructure and policy.

Summative scoring without weighting was used to enable simple comparisons of these measures across time and settings. Frequency of repeated incidents of forms of mistreatment was considered for weighting in the Interpersonal Abuse Scale since items in this measure could be based on multiple incident reports. However, accounting for frequency would embed a severity gradient in the scale scores that may not make sense conceptually and clinically. While PCA is a commonly used method to derive weights, this approach yields unique weights based on a specific set of data. It was not appropriate for measures developed from multicountry data and designed to be compared across multiple settings.^{63 64} Data-derived weights also would have removed women's experiences from the focus of mistreatment, placing an external valence on which mistreatment items are 'worse'

Table 3 Distribution of mistreatment items by measure and country (N=1974, unless otherwise noted)

Item	Nigeria (n=407) %	Ghana (n=912) %	Guinea (n=655) %
Interpersonal Abuse Scale (7 items)			
1. Shouted/screamed at	52.6	20.7	21.2
2. Insulted	20.2	2.9	2.8
3. Scolded	26.5	8.1	11.9
4. Mocked	21.4	2.4	7.9
5. Negative comments *	11.6	5.4	2.3
6. Threatened †	16.5	5.9	3.4
7. Physical abuse ‡	27.0	8.1	15.9
<i>Mean Scale Score (±SD)</i>	1.8 (1.8)	0.5 (1.0)	0.7 (1.0)
<i>Maximum score: 7 points</i>	(range: 0–7)	(range: 0–6)	(range: 0–7)
<i>% of women > country mean</i>	48.2	31.4	39.9
Exams & Procedures Index (3 items) (N=1538)			
1. No informed consent	78.6	72.5	50.1
2. Exposed	75.7	6.7	28.9
3. Confidential information	47.4	4.1	9.7
<i>Mean Index Score (±SD)</i>	2.0 (0.97)	0.8 (0.60)	0.9 (0.76)
<i>Maximum score: 3 points</i>	(range: 0–3)	(range: 0–3)	(range: 0–3)
<i>% of women > country mean</i>	38.9	74.3	66.8
Unsupportive Birth Environment Index (12 items)			
1. No pain relief	92.9	86.1	87.5
2. No interpreter	0.7	1.3	0.3
3. No staff present at birth	7.6	1.9	8.1
4. Not offered labour companion	96.1	93.6	88.4
5. Neglected	1.5	0.2	0.8
6. Bribe	3.4	0.7	6.4
7. Clean up blood/fluids	1.5	--	0.2
8. Fluids	43.0	43.6	30.8
9. Mobilise	93.6	75.9	19.1
10. No curtains/partitions	88.5	15.1	63.5
11. No bed	1.2	0.9	4.6
12. Shared bed	0.3	1.2	21.8
<i>Mean Index Score (±SD)</i>	4.3 (0.81)	3.2 (1.0)	3.3 (1.1)
<i>Maximum score: 12 points</i>	(range: 2–7)	(range: 1–6)	(range: 1–7)
<i>% of women > country mean</i>	40.1	42.0	39.5

*Assessed as: Did the woman receive negative comments about: (1) her physical appearance (including her weight, genitalia, cleanliness or other aspects of a woman's body), (2) her sexual activity, (3) or her baby's appearance (including his/her sex or other aspects of the baby)?

†Assessed as: Was the woman threatened with: (1) physical violence, (2) a poor outcome for her or her baby, (3) withholding care from her or her baby or (4) with the use of a medical procedure (such as episiotomy, Caesarean or other procedure)?

‡Assessed as: Did the woman experience at least one of the following?: pinched, kicked, slapped, punched or hit with instrument, gagged, physically tied to the bed, held down to the bed forcefully, given forceful downward pressure on abdomen, other use of physical force.

than others. This may be inconsistent across settings and individual experiences as societal norms, context, expectations and preferences inform what is viewed as mistreatment.

Results of psychometric analyses support a 7-item unidimensional Interpersonal Abuse Scale based on strong factor loadings of all seven items on a single factor in the three countries. While the scale captures multiple forms of verbal abuse, our assessment of physical abuse as

a single item due to low frequency of most forms of physical abuse may skew scale scores by types of verbal abuse and it may limit our ability to understand nuance in these experiences. The physical abuse item includes two dimensions: being physically struck (pinched, kicked, slapped, punched or hit with an instrument), and other uses of physical restraint or force (being gagged, physically tied to the bed, held down forcefully to the bed, forceful downward pressure on the abdomen and other use of

Table 4 Interpersonal Abuse Scale: results of exploratory factor analysis by country

Item (7 items)	Nigeria (N=407)		Ghana (N=912)		Guinea (N=655)	
	Factor loading	Uniqueness	Factor loading	Uniqueness	Factor loading	Uniqueness
1. Shouted/screamed at	0.77	0.41	0.84	0.29	0.68	0.54
2. Insulted	0.77	0.40	0.71	0.50	0.72	0.48
3. Scolded	0.72	0.48	0.48	0.77	0.58	0.67
4. Mocked	0.67	0.55	0.68	0.53	0.74	0.46
5. Negative comments	0.56	0.68	0.46	0.78	0.49	0.76
6. Threatened	0.64	0.60	0.69	0.53	0.50	0.75
7. Physical abuse *	0.58	0.67	0.64	0.60	0.55	0.70
Internal reliability						
KR-20 coefficient	0.71		0.57		0.54	
<i>Deleted items</i>						
1. Was the woman blamed Omitted due to low frequency for her or her baby's poor health outcomes?						
2. Was the woman hissed Nigeria-specific item at?						

*Any physical abuse=at least one experience of the following: pinched, kicked, slapped, punched, hit with instrument, gagged, physically tied to the bed, held down to the bed forcefully, given forceful downward pressure on abdomen, other use of physical force.
KR-20, Kuder-Richardson-20 coefficient.

force). Future validation of this scale should examine the validity of including these dimensions as separate scale items, as the impetus for these forms of abuse may have different underpinnings.^{13 49–51 53 74 75 81 82} While the low-frequency of stigma and discrimination items did not scale in our samples, we note that these critical forms of mistreatment may be more common and observable in other contexts, and their inclusion in future forms of the measures should be assessed.

Most women scored 1 or greater on the Exams& Procedures Index across all three countries, indicating that inappropriate conduct of exams and procedures, unconsented care and breaches of privacy were common. Breaches of privacy and exposing the woman's body may be influenced by the facility birth environment such as different physical layouts, space limitations and crowding in study facility. Sen *et al.* discussed normalised practices in low-income and middle-income country settings such as orienting beds and labour tables in open wards towards nurse stations and entryways to prioritise efficiency of labour progress monitoring over women's privacy, particularly when there is a limited number of providers.⁷² In contrast, unconsented care may be influenced by differences in norms or standards in clinical practice, perceived health literacy of the women, or power relations prioritising clinical decision making over two-way communication or women's preferences.

Some items in the Unsupportive Birth Environment Index refer specifically to the birth environment and facility resources (e.g., availability of fluids, curtains or partitions, no bed or bed sharing), and are similar to those captured in widely-used facility assessments like the

Service Provision Assessment (SPA) or the WHO Service Availability and Readiness Assessment (SARA). One important difference is that the mistreatment items are woman-centred in that they reflect how women interact with their birth environments. A facility may technically have clean water available or an adequate number of beds to accommodate their monthly volume, but this index measures whether a woman actually had access to and used these resources at an individual level. Two items in this index occurred in the vast majority of observations, no offer or denial of pain relief and not being offered a labour companion occurred in most observations, though they did not reach our criteria for omission of >90% frequency in all three samples. They were retained due to their theoretical importance in contributing to an unsupportive birth environment; both items signal areas for improvement on two evidence-based recommendations by the WHO for a positive birth experience.⁸³ Future refinement of the labour companion item in this index should consider using language used in the WHO community survey tool regarding whether a woman was 'allowed' to have a companion (instead of 'offered') as a targeted assessment of denying a woman's preferences for companionship.

Future validation of all three measures should further assess criterion validity given our mixed results from the predictive criterion analysis, as well confirmatory validation approaches and further assessment of measurement invariance. However, caution should be used when interpreting these findings, as it is difficult to determine whether the inconsistent findings are due to the actual validity of the measures or due to the criterion against

Table 5 Percent distribution of women experiencing a mistreatment item by mistreatment score and country*

Mistreatment item	Nigeria (N=407)		Ghana (N=912)		Guinea (N=655)	
	Low score	High score	Low score	High score	Low score	High score
Interpersonal Abuse Scale (7 items)						
Shouted	20.4	87.2	0.0	66.1	0.0	53.3
Insulted	2.8	38.8	0.0	9.1	0.0	6.9
Scolded	3.8	51.0	0.0	25.9	0.0	29.9
Mocked	1.4	42.9	0.0	7.7	0.0	19.9
Negative comments	0.5	23.5	0.0	17.1	0.0	5.8
Threatened	1.4	32.7	0.0	18.9	0.0	8.4
Physical abuse	6.6	49.0	0.0	25.9	0.0	39.9
Total women (n)	211	196	626	286	394	261
<i>Country mean score (unweighted) (SD)</i>	1.8 (1.8) (range: 0–7)		0.54 (1.0) (range: 0–6)		0.65 (1.0) (range: 0–7)	
<i>% above country mean ('high')</i>	48.2		31.4		39.9	
Exams & Procedures Index (3 items)	(n=350)		(n=797)		(n=391)	
Informed consent	65.0	100.0	0.0	97.6	0.0	56.7
Exposed	60.3	100.0	0.0	9.0	0.0	43.3
Confidential information	14.0	100.0	0.0	5.6	0.0	14.6
Total Women (n)	214	136	205	592	394	261
<i>Country mean score (unweighted) (SD)</i>	2.0 (0.97) (range: 0–3)		0.83 (0.60) (range: 0–3)		0.89 (0.78) (range: 0–3)	
<i>% above country mean ('high')</i>	38.9		74.3		66.8	
Unsupportive Birth Environment Index (12 items)						
Pain relief	88.9	98.8	76.8	99.0	81.6	96.5
No interpreter	0.0	1.8	0.6	2.4	0.0	0.8
No staff present at birth	4.9	11.7	0.6	3.7	3.8	14.7
Not offered labour companion	93.9	99.4	89.6	99.2	85.4	93.1
Neglect	0.0	3.7	0.2	0.3	0.0	1.9
Bribe	1.2	6.8	0.6	0.8	1.0	14.7
Clean up blood/fluids	0.4	3.1	--	--	0.0	0.4
Fluids	13.1	87.7	14.2	84.3	14.4	56.0
Mobilise	89.8	99.4	59.9	97.9	5.8	39.4
No curtains/partitions	82.8	96.9	5.5	28.5	50.5	83.4
No bed	0.4	2.5	0.8	1.0	0.5	10.8
Shared bed	0.0	0.6	0.0	2.9	11.6	37.5
Total women (n)	244	163	529	383	396	259
<i>Country mean score (unweighted) (SD)</i>	4.3 (0.81) (range: 2–7)		3.2 (1.0) (range: 1–6)		3.3 (1.1) (range: 1–7)	
<i>% above country mean ('high')</i>	40.1		42.0		39.5	

*Table shows an evaluation of the consistency of scores for each measure. Proportions shown compare row frequencies of women in 'high' and 'low' scoring groups with an observed mistreatment item (determined by those scoring at/above or below the country-specific mean score on the measure). Internally consistent scores have a higher proportion of women in the 'high' scoring group experiencing each item. Frequencies are comparative and do not sum to 100.

which they were measured (or both). We used global measures of satisfaction with care and future intentions to deliver in the same facility to be consistent with studies that have used this criterion to validate composite

measures of respectful maternity care³⁵ and person-centred maternity care.^{41 42} However, the extent to which women's reports of satisfaction are consistently related to mistreatment, particularly third-party labour observations

of interpersonal abuse, inappropriate conduct of exams and procedures or an unsupportive birth environment is unclear. A recent analysis of the larger WHO study community survey data found that while over one-third of women experienced mistreatment, nearly 90% of women reported overall satisfaction with care received.⁸⁴ These results highlight the complexity of comprehensively measuring quality of care given this duality of high satisfaction as an outcome measure—informed by expectations of care, empowerment, and knowledge of rights to high quality and respectful care—in the presence of mistreatment as a discrete process measure of women's experiences with health systems.^{9 84} Taken together, our confidence in relying on these criteria to help firmly establish validity of these measures is limited. As research in this field evolves and more measures are developed and refined, further validation of our measures of care quality and experience can be tested against other mistreatment measures to help determine their criterion validity. Further refinement and expansion of items in the indexes may also be needed to iteratively maintain content validity, as the content variety and scope of items is particularly important for the validity of indexes.

These measures were developed using data from three West African countries; their application in other geographical settings and types of facilities (health centres, district hospitals, tertiary hospitals) will also help determine their construct validity and generalisability. While the measures showed good internal reliability, further reliability testing of the Interpersonal Abuse Scale in other settings is needed as the reliability coefficient in Ghana and Guinea did not provide strong evidence of scale reliability. Inter-rater reliability assessments were not possible with this data, but should also be conducted in future studies to further establish reliability of the measures, particularly since they are based on labour observations.

Strengths and limitations

A major strength of this study is the use of data from a multicountry study, adding to the robustness of the validation of the newly developed measures. A strength of the use of continuous labour observations as the basis of these measures is that they limit potential recall and social desirability biases that can accompany women's reports of mistreatment. Continuous observations and recording of mistreatment in real time is particularly important to capture mistreatment during birth, as the primary WHO study analysis found mistreatment peaked around the time of birth.⁸⁵

There are several limitations of this analysis. All study facilities were public and located in urban settings, limiting generalisability of the mistreatment measures given previous research has found differences in how women are treated during childbirth by type of facility.^{15 21 22 42 86 87} Limited information was available in our data on technical and human resources available in facilities as well as more detailed information

on providers such as gender and experience levels; this information may be useful to further assess the construct validity based on prior research that shows different likelihood of mistreatment by various provider characteristics.^{11 20 88} Assessing items in the Interpersonal Abuse Scale as binary rather than incorporating repeated incidents could have reduced data variability. Using continuous observations poses a potential Hawthorn effect, though previously conducted assessments of this data did not show a significant presence of this effect.¹⁰ Labour observations may not be the most accurate mode to assess more experiential aspects of mistreatment (i.e., aspects of mistreatment that may be intrinsically based on women's perceptions on care experienced, such as stigma or discrimination).⁹ Our measures may include items that could be readily observed but 'matter' less to women's personal experience of care, removing items that may be important to a women's birth experience such as blaming women for poor outcomes. Triangulating labour observation measures with measures based on women's reports of mistreatment, like those developed in an analysis of the WHO community survey data⁸⁹, aligns multiple perspectives to enhance our understanding of rigorously measuring this complex phenomenon. Other methodological limitations using the study dataset and tools have been previously described.¹⁰

Implications for policy and practice

These measures can be adapted and used in future research on mistreatment to quantify the burden, frequency and overlap in multiple types of mistreatment in a standardised way that can be compared across studies, settings and over time.

The concise nature of the measures, compared to the 56-item WHO labour observation mistreatment instrument, offers an opportunity to incorporate them in longer surveys of other aspects of maternal health or for quality improvement initiatives. The measures can also be used in conjunction with facility assessment tools like the SPA or SARA to gain a better understanding of the relationship between the health system or facility context and dimensions of mistreatment. The elements of interpersonal behaviour and abuse as well as health system infrastructure and resource constraints in the measures allow for their use to assess multicomponent interventions that impact both interpersonal aspects of care and more systemic, structural aspects of the process of care. The measures can also translate evidence-based recommendations into indicators that can be used for routine measurement of mistreatment, quality assessments and monitoring and accountability efforts to improve quality of maternal healthcare at the country level.^{90–92} This kind of routine measurement is essential to monitor progress towards the promotion of high-quality, respectful birth experiences for all women.

Author affiliations

¹Population, Family and Reproductive Health, Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland, USA

²Department of Sexual and Reproductive Health and Research, including UNDP/UNFPA/UNICEF/WHO/World Bank Special Programme of Research, Development and Research Training in Human Reproduction (HRP), World Health Organization, Geneva, Switzerland

³Centre for Health Equity, University of Melbourne School of Population and Global Health, Melbourne, Victoria, Australia

⁴Department of Obstetrics and Gynaecology, School of Medicine and Dentistry, University of Ghana, Accra, Greater Accra, Ghana

⁵Division of Prevention Science, University of California, San Francisco, San Francisco, California, USA

⁶Obstetrics and Gynaecology, University of Medical Sciences Teaching Hospital, Akure, Ondo, Nigeria

⁷Department of Medical Research, Ministry of Health and Sports, Yangon, Myanmar

⁸Cellule de Recherche en Sante de la Reproduction en Guinee (CERREGUI), University National Hospital-Donka, Conakry, Guinea

⁹UNDP/UNFPA/ UNICEF/WHO/World Bank Special Programme of Research, Development and Research Training in Human Reproduction (HRP), Department of Sexual and Reproductive Health and Research, World Health Organization, Geneva, Switzerland

Twitter Blair O Berger @blair_berger, Hedieh Mehrtash @hediehhm and Özge Tunçalp @otuncalp

Acknowledgements We thank the data collection team for their excellent work and the women who participated in this study. We appreciate the thoughtful contributions of participants in the end-of-study investigator's meeting. We also thank Dr. Karen Bandeen-Roche (Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health) for her technical guidance during the conceptualisation and conduct of this analysis.

Contributors BOB led the analysis, drafted the manuscript, and read and approved the final manuscript. DMS contributed to the analysis and drafting of the manuscript, and read and approved the final manuscript. HM provided input on the analysis and drafting of the manuscript, and read and approved the final manuscript. MAB provided input on the analysis and read and approved the final manuscript. KA-B, HLL, TAI, TMM and MDB provided input on the drafting of the manuscript and read and approved the final manuscript. OT provided input on the analysis and drafting of the manuscript, and read and approved the final manuscript.

Funding This research was made possible by the support of the American People through the United States Agency for International Development (USAID), and the UNDP/UNFPA/UNICEF/WHO/World Bank Special Programme of Research, Development and Research Training in Human Reproduction (HRP), Department of Reproductive Health and Research, WHO.

Disclaimer The contents of this article are the sole responsibility of the authors and do not necessarily reflect the views of USAID, the United States Government, WHO, or their individual institutions.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval Institutional permission for recruitment and observation was obtained from each site; consent was not sought from providers. This study was approved by the WHO Ethical Review Committee (A65880), WHO Review Panel on Research Projects, and in-country ethics committees: Guinea (le comité national d'éthique pour la recherche en santé); Nigeria [Federal Capital Territory Health Research Ethics Committee; Research Ethical Review Committee, Oyo State; and State Health Research Ethics Committee of Ondo State]; and Ghana (Ethical Review Committee of the Ghana Health Service; Ethical and Protocol Review Committee of the College of Health Sciences, University of Ghana), and Myanmar (Ethics Review Committee, Department of Medical Research).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on request. The analytical study dataset from the 'How women are treated during facility-based childbirth' WHO study is deidentified and, archived through WHO/HRP's electronic recordmanagement system. Data requests with an expression of interest in pursuing multicountrysecondary analyses with a specific research question can be made to srhmp@who.int. More information about the study tools are available here: <https://bmcmredresmethodol.biomedcentral.com/articles/10.1186/s12874-018-0603-x> and the primary publication from the study here: [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(19\)31992-0/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(19)31992-0/fulltext).

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Blair O Berger <http://orcid.org/0000-0002-7962-0522>

Meghan A Bohren <http://orcid.org/0000-0002-4179-4682>

Hannah H Leslie <http://orcid.org/0000-0002-7464-3645>

Thae Maung Maung <http://orcid.org/0000-0002-1265-3813>

REFERENCES

- Bohren MA, Hunter EC, Munthe-Kaas HM, *et al*. Facilitators and barriers to facility-based delivery in low- and middle-income countries: a qualitative evidence synthesis. *Reprod Health* 2014;11:71.
- Miller S, Abalos E, Chamillard M, *et al*. Beyond too little, too late and too much, too soon: a pathway towards evidence-based, respectful maternity care worldwide. *Lancet* 2016;388:2176–92.
- Gabrysch S, Campbell OMR. Still too far to walk: literature review of the determinants of delivery service use. *BMC Pregnancy Childbirth* 2009;9:34.
- Souza JP, Gülmezoglu AM, Vogel J, *et al*. Moving beyond essential interventions for reduction of maternal mortality (the who multicountry survey on maternal and newborn health): a cross-sectional study. *Lancet* 2013;381:1747–55.
- Hulton L, Matthews Z, Martin-Hilber A, *et al*. Using evidence to drive action: a "revolution in accountability" to implement quality care for better maternal and newborn health in Africa. *Int J Gynaecol Obstet* 2014;127:96–101.
- Ö T, Were W, MacLennan C. Quality of care for pregnant women and newborns—the who vision. *BJOG: an international journal of obstetrics & gynaecology* 2015;122:1045–9.
- Brizuela V, Leslie HH, Sharma J, *et al*. Measuring quality of care for all women and newborns: how do we know if we are doing it right? A review of facility assessment tools. *Lancet Glob Health* 2019;7:e624–32.
- Kruk ME, Gage AD, Arsenault C, *et al*. High-Quality health systems in the sustainable development goals era: time for a revolution. *Lancet Glob Health* 2018;6:e1196–252.
- Larson E, Sharma J, Bohren MA. *When the patient is the expert: measuring patient experience and satisfaction with care*. Bulletin of the World Health Organization, 2019: 1–13.
- Bohren MA, Mehrtash H, Fawole B, *et al*. How women are treated during facility-based childbirth in four countries: a cross-sectional study with labour observations and community-based surveys. *Lancet* 2019;394:1750–63.
- Abuya T, Warren CE, Miller N, *et al*. Exploring the prevalence of disrespect and abuse during childbirth in Kenya. *PLoS One* 2015;10:e0123606.
- Bohren MA, Vogel JP, Hunter EC, *et al*. The mistreatment of women during childbirth in health facilities globally: a mixed-methods systematic review. *PLoS Med* 2015;12:e1001847.
- Rominski SD, Lori J, Nakua E, *et al*. When the baby remains there for a long time, it is going to die so you have to hit her small for the baby to come out": justification of disrespectful and abusive care during childbirth among midwifery students in Ghana. *Health Policy Plan* 2017;32:215–24.
- Bowser D, Hill K. *Exploring evidence for disrespect and abuse in facility-based childbirth*. Boston: USAID-TRAction Project, Harvard School of Public Health, 2010.
- Sethi R, Gupta S, Oseni L, *et al*. The prevalence of disrespect and abuse during facility-based maternity care in Malawi: evidence from direct observations of labor and delivery. *Reprod Health* 2017;14:111.

- 16 Sando D, Ratcliffe H, McDonald K, *et al*. The prevalence of disrespect and abuse during facility-based childbirth in urban Tanzania. *BMC Pregnancy Childbirth* 2016;16:236.
- 17 Raj A, Dey A, Boyce S, *et al*. Associations between mistreatment by a provider during childbirth and maternal health complications in Uttar Pradesh, India. *Matern Child Health J* 2017;21:1821–33.
- 18 Okafor II, Ugwu EO, Obi SN. Disrespect and abuse during facility-based childbirth in a low-income country. *Int J Gynaecol Obstet* 2015;128:110–3.
- 19 Kujawski S, Mbaruku G, Freedman LP, *et al*. Association between Disrespect and abuse during childbirth and women's confidence in health facilities in Tanzania. *Matern Child Health J* 2015;19:2243–50.
- 20 Sheferaw ED, Bazant E, Gibson H, *et al*. Respectful maternity care in Ethiopian public health facilities. *Reprod Health* 2017;14:60.
- 21 Abuya T, Ripad P, Ritter J, *et al*. Measuring mistreatment of women throughout the birthing process: implications for quality of care assessments. *Reprod Health Matters* 2018;26:48–61.
- 22 Afulani PA, Phillips B, Aborigo RA, *et al*. Person-Centred maternity care in low-income and middle-income countries: analysis of data from Kenya, Ghana, and India. *Lancet Glob Health* 2019;7:e96–109.
- 23 Bishanga DR, Massenga J, Mwanamsangu AH, *et al*. Women's experience of Facility-Based childbirth care and receipt of an early postnatal check for herself and her newborn in northwestern Tanzania. *Int J Environ Res Public Health* 2019;16. doi:10.3390/ijerph16030481. [Epub ahead of print: 07 Feb 2019].
- 24 Tekle Bobo F, Kebebe Kasaye H, Etana B, *et al*. Disrespect and abuse during childbirth in Western Ethiopia: should women continue to tolerate? *PLoS One* 2019;14:e0217126.
- 25 Ijadunola MY, Olotu EA, Oyedun OO, *et al*. Lifting the veil on disrespect and abuse in facility-based child birth care: findings from South West Nigeria. *BMC Pregnancy Childbirth* 2019;19:39.
- 26 Vedam S, Stoll K, McRae DN, *et al*. Patient-led decision making: measuring autonomy and respect in Canadian maternity care. *Patient Educ Couns* 2019;102:586–94.
- 27 Morton CH, Henley MM, Seacrist M, *et al*. Bearing witness: United States and Canadian maternity support workers' observations of disrespectful care in childbirth. *Birth* 2018;45:263–74.
- 28 Sando D, Abuya T, Asefa A, *et al*. Methods used in prevalence studies of disrespect and abuse during facility based childbirth: lessons learned. *Reprod Health* 2017;14:127.
- 29 Sharma G, Penn-Kekana L, Halder K, *et al*. An investigation into mistreatment of women during labour and childbirth in maternity care facilities in Uttar Pradesh, India: a mixed methods study. *Reprod Health* 2019;16:7.
- 30 Warren C, Njuki R, Abuya T, *et al*. Study protocol for promoting respectful maternity care initiative to assess, measure and design interventions to reduce disrespect and abuse during childbirth in Kenya. *BMC Pregnancy Childbirth* 2013;13:21.
- 31 Abuya T, Ndwiga C, Ritter J, *et al*. The effect of a multi-component intervention on disrespect and abuse during childbirth in Kenya. *BMC Pregnancy Childbirth* 2015;15:224.
- 32 Kruk ME, Kujawski S, Mbaruku G. Disrespectful and abusive treatment during facility delivery in Tanzania: a facility and community survey. *Health Policy Plan* 2014;czu079.
- 33 Rosen HE, Lynam PF, Carr C, *et al*. Direct observation of respectful maternity care in five countries: a cross-sectional study of health facilities in East and southern Africa. *BMC Pregnancy Childbirth* 2015;15:306.
- 34 Sando D, Kendall T, Lyatuu G, *et al*. Disrespect and abuse during childbirth in Tanzania: are women living with HIV more vulnerable? *J Acquir Immune Defic Syndr* 2014;67 Suppl 4:S228.
- 35 Sheferaw ED, Mengesha TZ, Wase SB. Development of a tool to measure women's perception of respectful maternity care in public health facilities. *BMC Pregnancy Childbirth* 2016;16:67.
- 36 Ansara DL, Hindin MJ. Exploring gender differences in the patterns of intimate partner violence in Canada: a latent class approach. *J Epidemiol Community Health* 2010;64:849–54.
- 37 Straus MA. Measuring intrafamily conflict and violence: the conflict tactics (CT) scales. *J Marriage Fam* 1979;41:75–88.
- 38 Straus MA, Hamby SL, Boney-McCoy S. The revised conflict tactics scales (CTS2) development and preliminary psychometric data. *J Fam Issues* 1996;17:283–316.
- 39 Hindin MJ. Understanding women's attitudes towards wife beating in Zimbabwe. *Bull World Health Organ* 2003;81:501–8.
- 40 Ellsberg M, Heise L, World Health Organization. *Researching violence against women: a practical guide for researchers and activists*, 2005.
- 41 Afulani PA, Diamond-Smith N, Golub G, *et al*. Development of a tool to measure person-centered maternity care in developing settings: validation in a rural and urban Kenyan population. *Reprod Health* 2017;14:118.
- 42 Afulani PA, Diamond-Smith N, Phillips B, *et al*. Validation of the person-centered maternity care scale in India. *Reprod Health* 2018;15:147.
- 43 Afulani PA, Feeser K, Sudhinaraset M, *et al*. Toward the development of a short multi-country person-centered maternity care scale. *Int J Gynaecol Obstet* 2019;146:80–7.
- 44 Dey A, Shakya HB, Chandurkar D, *et al*. Discordance in self-report and observation data on mistreatment of women by providers during childbirth in Uttar Pradesh, India. *Reprod Health* 2017;14:149.
- 45 Diamond-Smith N, Treleaven E, Murthy N, *et al*. Women's empowerment and experiences of mistreatment during childbirth in facilities in Lucknow, India: results from a cross-sectional study. *BMC Pregnancy Childbirth* 2017;17:335.
- 46 Blanc AK, Warren C, McCarthy KJ, *et al*. Assessing the validity of indicators of the quality of maternal and newborn health care in Kenya. *J Glob Health* 2016;6:010405.
- 47 Bohren MA, Vogel JP, Fawole B, *et al*. Methodological development of tools to measure how women are treated during facility-based childbirth in four countries: labor observation and community survey. *BMC Med Res Methodol* 2018;18:132.
- 48 Vogel JP, Bohren MA, Tunçalp Özge, *et al*. How women are treated during facility-based childbirth: development and validation of measurement tools in four countries - phase 1 formative research study protocol. *Reprod Health* 2015;12:60.
- 49 Maya ET, Adu-Bonsaffoh K, Dako-Gyeke P, *et al*. Women's perspectives of mistreatment during childbirth at health facilities in Ghana: findings from a qualitative study. *Reprod Health Matters* 2018;26:70–87.
- 50 Balde MD, Bangoura A, Diallo BA, Sall O, *et al*. A qualitative study of women's and health providers' attitudes and acceptability of mistreatment during childbirth in health facilities in guinea. *Reprod Health* 2017;14:4.
- 51 Balde MD, Diallo BA, Bangoura A, *et al*. Perceptions and experiences of the mistreatment of women during childbirth in health facilities in guinea: a qualitative study with women and service providers. *Reprod Health* 2017;14:3.
- 52 Bohren MA, Vogel JP, Tunçalp Özge, *et al*. Mistreatment of women during childbirth in Abuja, Nigeria: a qualitative study on perceptions and experiences of women and healthcare providers. *Reprod Health* 2017;14:9.
- 53 Bohren MA, Vogel JP, Tunçalp Özge, *et al*. "By slapping their laps, the patient will know that you truly care for her": A qualitative study on social norms and acceptability of the mistreatment of women during childbirth in Abuja, Nigeria. *SSM Popul Health* 2016;2:640–55.
- 54 Maung TM, Show KL, Mon NO, *et al*. A qualitative study on acceptability of the mistreatment of women during childbirth in Myanmar. *Reprod Health* 2020;17:56.
- 55 StataCorp LP. *Stata/MP 13.0. College Station, tx*, 2013.
- 56 Bollen K, Lennox R. Conventional wisdom on measurement: a structural equation perspective. *Psychol Bull* 1991;110:305–14.
- 57 Streiner DL. Being inconsistent about consistency: when coefficient alpha does and doesn't matter. *J Pers Assess* 2003;80:217–22.
- 58 DeVellis RF. *Scale development: theory and applications*. Sage Publications, 2016.
- 59 Netemeyer RG, Bearden WO, Sharma S. *Scaling procedures: issues and applications*. Sage Publications, 2003.
- 60 Clark LA, Watson D. Constructing validity: basic issues in objective scale development. *Psychol Assess* 1995;7:309–19.
- 61 Mokkink LB, Prinsen C, Patrick DL. COSMIN methodology for systematic reviews of patient-reported outcome measures (PROMs). *User manual* 2018;78:1.
- 62 Prinsen CAC, Vohra S, Rose MR, *et al*. How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" - a practical guideline. *Trials* 2016;17:449.
- 63 OECD. *Handbook on constructing composite indicators: methodology and user guide*. OECD publishing, 2008.
- 64 Detrick Z, Gouda HN, Hodge A, *et al*. Measuring quality of maternal and newborn care in developing countries using demographic and health surveys. *PLoS One* 2016;11:e0157110.
- 65 Tripathi V, Stanton C, Strobino D, *et al*. Development and validation of an index to measure the quality of Facility-Based labor and delivery care processes in sub-Saharan Africa. *PLoS One* 2015;10:e0129491.
- 66 Wilhelm D, Lohmann J, De Allegri M, *et al*. Quality of maternal obstetric and neonatal care in low-income countries: development of a composite index. *BMC Med Res Methodol* 2019;19:154.
- 67 Pirkle CM, Dumont A, Traore M, *et al*. Validity and reliability of criterion based clinical audit to assess obstetrical quality of care in West Africa. *BMC Pregnancy Childbirth* 2012;12:118.
- 68 Jennings L, Na M, Cherewick M, *et al*. Women's empowerment and male involvement in antenatal care: analyses of demographic and

- health surveys (DHS) in selected African countries. *BMC Pregnancy Childbirth* 2014;14:297.
- 69 Coltman T, Devinney TM, Midgley DF, *et al*. Formative versus reflective measurement models: two applications of formative measurement. *J Bus Res* 2008;61:1250–62.
- 70 Bollen KA, Davis WR. Causal indicator models: identification, estimation, and testing. *Structural Equation Modeling: A Multidisciplinary Journal* 2009;16:498–522.
- 71 Strobino DM, Baruffi G. Evaluation of a measure of neonatal morbidity. *Med Care* 1984;22:818–26.
- 72 Sen G, Reddy B, Iyer A. Beyond measurement: the drivers of disrespect and abuse in obstetric care. *Reprod Health Matters* 2018;26:6–18.
- 73 Moyer CA, Rominski S, Nakua EK, *et al*. Exposure to disrespectful patient care during training: data from midwifery students at 15 midwifery schools in Ghana. *Midwifery* 2016;41:39–44.
- 74 Jewkes R, Abrahams N, Mvo Z. Why do nurses abuse patients? reflections from South African obstetric services. *Soc Sci Med* 1998;47:1781–95.
- 75 Kruger Lou-Marié, Schoombee C. The other side of caring: abuse in a South African maternity ward. *J Reprod Infant Psychol* 2010;28:84–101.
- 76 Smith PH, Earp JA, DeVellis R. Measuring battering: development of the Women's Experience with Battering (WEB) Scale. *Women's Health* 1995;1:273–88.
- 77 Baba K, Takauma F, Tada K, *et al*. Factor structure of the conflict tactics scale 1. *Int J Community Based Nurs Midwifery* 2017;5:239–47.
- 78 Kalokhe AS, Stephenson R, Kelley ME, *et al*. The development and validation of the Indian family violence and control scale. *PLoS One* 2016;11:e0148120.
- 79 Marshall LL. Development of the severity of violence against women scales. *J Fam Violence* 1992;7:103–21.
- 80 de Vet HCW, Terwee CB, Bouter LM. Current challenges in clinimetrics. *J Clin Epidemiol* 2003;56:1137–41.
- 81 Burrowes S, Holcombe SJ, Jara D, *et al*. Midwives' and patients' perspectives on disrespect and abuse during labor and delivery care in Ethiopia: a qualitative study. *BMC Pregnancy Childbirth* 2017;17:263.
- 82 Warren N, Beebe M, Chase RP, *et al*. Nègèngè: sweet talk, disrespect, and abuse among rural auxiliary midwives in Mali. *Midwifery* 2015;31:1073–80.
- 83 World Health Organization. *Who recommendations: intrapartum care for a positive childbirth experience*. World Health Organization, 2018.
- 84 Maung TM, Mon NO, Mehrtash H, *et al*. Women's experiences of mistreatment during childbirth and their satisfaction with care: findings from a multicountry community-based study in four countries. *BMJ Glob Health* 2021;5:e003688.
- 85 Bohren MA, Berger BO, Munthe-Kaas H, *et al*. Perceptions and experiences of labour companionship: a qualitative evidence synthesis. *Cochrane Database Syst Rev* 2019;3:CD012449.
- 86 Afulani PA, Sayi TS, Montagu D. Predictors of person-centered maternity care: the role of socioeconomic status, empowerment, and facility type. *BMC Health Serv Res* 2018;18:360.
- 87 Asefa A, Bekele D. Status of respectful and non-abusive care during facility-based childbirth in a hospital and health centers in Addis Ababa, Ethiopia. *Reprod Health* 2015;12:33.
- 88 Jewkes R, Penn-Kekana L. Mistreatment of women in childbirth: time for action on this important dimension of violence against women. *PLoS Med* 2015;12:e1001849.
- 89 Leslie HH, Sharma J, Mehrtash H. Women's report of mistreatment during facility-based childbirth: validity and reliability of community survey measures. *BMJ Global Health* 2021;5:e004822.
- 90 Afulani PA, Buback L, McNally B, *et al*. A rapid review of available evidence to inform indicators for routine monitoring and evaluation of Respectful maternity care. *Glob Health Sci Pract* 2020;8:125–35.
- 91 Adeniran A, Likaka A, Knutsson A, *et al*. Leadership, action, learning and accountability to deliver quality care for women, newborns and children. *Bull World Health Organ* 2018;96:222–4.
- 92 World Health Organization. *Quality, equity, dignity: the network to improve quality of care for maternal, newborn and child health: strategic objectives*, 2018.

Appendix 1. Analytic Approach

1. Item Construction and Identifying Dimensions

1a. Exploratory data analysis: Item frequencies were calculated, and items were combined based on theory, practicality, understandability and item translation across study countries. Patterns of repeated events and co-occurring mistreatment incidents also were examined to determine the most appropriate scoring structure.

1b. To determine whether a psychometric approach was appropriate to develop the measures, a series of preliminary tests (examinations of principal components analysis (PCA), Bartlett's test of sphericity, the Kaiser-Meyer-Olin (KMO) measure of sampling adequacy) were performed on the dimensions separately and as a single composite measure to determine whether the items yielded a consistent factor structure (i.e., "factorability") (1, 2). To confirm this conclusion, PCA and exploratory factor analysis (EFA) were conducted on items from all dimensions, both separately by dimension and combined. *Results*: The final 7 scale items were satisfactory for factor analysis in all settings based on a significant p-value ≤ 0.05 on Bartlett's test of sphericity and Kaiser-Meyer-Olin (KMO) measures of sampling adequacy > 0.50 overall and for each item; the overall KMO was 0.75 in Nigeria, 0.61 in Ghana, and 0.76 in Guinea.

1c. Missing data: Completeness of the labor observation data was high for the mistreatment items at 98% (1,974/2,016). The pattern of missing data for the 42 women with incomplete or missing data collection forms was assessed within each country sample to determine if items were clustered by facility or by observer within a facility. While nearly two thirds of missing or incomplete data forms were in Guinea (27/42), no such clustering was observed. Women with missing data on the mistreatment items were not included in the analyses.

2. Interpersonal Abuse Scale Development: Psychometric Analysis

2a. Tetrachoric correlation matrices were examined during EFA to assess interrelatedness of the binary interpersonal abuse items in each country as well as to examine similarities and deviations in the factor structures across the three samples. Item uniqueness was assessed with the goal of minimizing uniqueness (using the benchmark of < 0.50 as "acceptable,"), though retaining items with high factor loading and theoretical importance to the construct were prioritized over low uniqueness (1). *Results*: While 3-4 items in each country had uniqueness values above the recommend value of 0.50, they were retained due to high factor loadings and theoretical importance of the items to the interpersonal abuse dimension.

2b. To assess reliability/internal consistency of the scale, a range of 0.15-0.50 corrected item-to-total and interitem correlations was used as evidence of internal consistency for new measures in early scale development (1-3). Scale reliability was also assessed via the Kuder-Richardson-20 (KR-20) coefficient (the corollary to Cronbach's alpha used for scales based on dichotomous items), where a coefficient of ≥ 0.70 is considered the benchmark for consistency in scale development (1, 2, 4, 5). *Results*: Tetrachoric correlations ranged between 0.22-0.62 in Nigeria, 0.24-0.58 in Ghana, and 0.18-0.57 in Guinea, with two exceptions: between threatened and insulted in Ghana and between physical abuse and negative comments in Guinea, both of which had pairwise correlations < 0.10 (**Appendix 1c**).

3. Index Development for Exams & Procedures and Unsupportive Birth Environment

3a. Consistent with the OECD approach for composite index development, item selection occurred using a theoretical framework (WHO typology of mistreatment), literature review, and expert opinion, and assessment of missing data and imputation during preliminary data analysis of the item pool (6).

3b. No normalization was required because all items were binary, defined as "0=no mistreatment" and "1=mistreatment."

4. Measure Validation and Reliability Assessment

4a. In assessing Spearman-Rank correlations between the three measures, correlations (<0.30 , $p>0.05$) across different measures provide evidence of construct validity and confirm that the measures should be assessed separately rather than as an overall composite measure.

4b. Predictive criterion validity (the extent to which measures are associated with theoretically related measures or outcomes in predictable ways) was assessed by regressing a global measure of satisfaction with care and intentions to delivery in the same facility in the future on ordinal mistreatment measure scores. These global satisfaction with care variables were obtained from linked data of women's responses on the WHO community survey tool. They have been used to assess construct validity in previous studies assessing mistreatment, RMC, and person-centered maternity care (7-12). We hypothesized that women with higher mistreatment scores would have: 1) higher odds of reporting overall dissatisfaction with care, and 2) lower odds of intending to deliver in the same facility for a future delivery. *Results:* Higher scale scores were positively associated with women's reports of dissatisfaction with their overall care experience in Ghana and Guinea, indicating predictive validity of the scale, though no association was observed in Nigeria. Evidence for predictive validity was limited for the Exams & Procedures Index and the Unsupportive Birth Environment Index (**Appendix 2b**).

4c. Reliability: Two additional tests of internal consistency were conducted on all three mistreatment measures to further determine the reliability of the Interpersonal Abuse Scale and to establish a degree of consistency in the two indexes. *To assess consistency of items within a measure:* the percent distributions of co-occurring mistreatment items were examined, whereby internally consistent measures would have higher proportions of items that occur with at least one or several other items in a measure (results available upon request). *To determine consistency of measure scores:* to determine if each item could distinguish between "high" and "low" scores of mistreatment, a binary variable was constructed where "high" scores included observations with scores higher than the country-specific mean, and "low" included those scoring at or below the country-specific mean (13, 14). The proportion of women experiencing a mistreatment item who had "high" and "low" scores was then calculated; if scores were internally consistent and distinguished between these two groups, we would expect to see a higher proportion of women in the "high" score group experiencing each item (14).

5. Performance of Measures

5a. Item frequencies, mean scores, and percent distributions of mistreatment scores were calculated separately by mistreatment measure and country.

5b. Other descriptive statistics were calculated to assess measure performance, including low and high scores and the range of item frequency (the number of items occurring in $<5\%$ and $>90\%$ of observations).

5c. Ceiling and floor effects were assessed through the percent distribution of observations with minimum and maximum scores, using 15% in either category as evidence of ceiling or floor effects; presence of these effects could indicate inadequate content validity (15, 16).

Selected statistics assessing full score distributions and overall measure performance are shown in **Appendix 4**.

6. Sensitivity Analyses

6a. Mistreatment items that included "don't know" or "not applicable" options that were originally coded as "0=no mistreatment" on the items to yield the most conservative estimates were recoded as "1=mistreatment" to yield the most inflated mistreatment estimates.

6b. Responses in the "don't know" or "not applicable" categories were also omitted to determine the impact of these structural changes on item inclusion in the final measures.

6c. Versions of the two indexes that included and omitted combinations of high and low frequency items were assessed for measure performance (based on criteria in Step 5).

6d. Measure score distributions and performance assessments and correlations between measures were calculated at the facility-level, rather than the country-level, to assess the level of heterogeneity in item and measure performance by facility.

References

1. DeVellis RF. Scale development: Theory and applications: Sage Publications; 2016.
2. Netemeyer RG, Bearden WO, Sharma S. Scaling procedures: Issues and applications: Sage Publications; 2003.
3. Clark LA, Watson D. Constructing validity: Basic issues in objective scale development. *Psychological assessment*. 1995;7(3):309.
4. Straus MA. Measuring intrafamily conflict and violence: The conflict tactics (CT) scales. *Journal of Marriage and the Family*. 1979;75-88.
5. Straus MA, Hamby SL, Boney-McCoy S, Sugarman DB. The revised conflict tactics scales (CTS2) development and preliminary psychometric data. *Journal of family issues*. 1996;17(3):283-316.
6. OECD. Handbook on constructing composite indicators: methodology and user guide: OECD publishing; 2008.
7. Afulani PA, Diamond-Smith N, Golub G, Sudhinaraset M. Development of a tool to measure person-centered maternity care in developing settings: validation in a rural and urban Kenyan population. *Reprod Health*. 2017;14(1):118.
8. Afulani PA, Diamond-Smith N, Phillips B, Singhal S, Sudhinaraset M. Validation of the person-centered maternity care scale in India. *Reprod Health*. 2018;15(1):147.
9. Afulani PA, Feeser K, Sudhinaraset M, Aborigo R, Montagu D, Chakraborty N. Toward the development of a short multi-country person-centered maternity care scale. *Int J Gynaecol Obstet*. 2019;146(1):80-7.
10. Vedam S, Stoll K, McRae DN, Korchinski M, Velasquez R, Wang J, et al. Patient-led decision making: Measuring autonomy and respect in Canadian maternity care. *Patient Educ Couns*. 2019;102(3):586-94.
11. Vedam S, Stoll K, Taiwo TK, Rubashkin N, Cheyney M, Strauss N, et al. The Giving Voice to Mothers study: inequity and mistreatment during pregnancy and childbirth in the United States. *Reprod Health*. 2019;16(1):77.
12. Sheferaw ED, Mengesha TZ, Wase SB. Development of a tool to measure women's perception of respectful maternity care in public health facilities. *BMC Pregnancy and Childbirth*. 2016;16(1):67.
13. Jennings L, Na M, Cherewick M, Hindin M, Mullany B, Ahmed S. Women's empowerment and male involvement in antenatal care: analyses of Demographic and Health Surveys (DHS) in selected African countries. *BMC Pregnancy Childbirth*. 2014;14:297.
14. Strobino DM, Baruffi G. Evaluation of a measure of neonatal morbidity. *Medical Care*. 1984;818-26.
15. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60(1):34-42.
16. Tripathi V, Stanton C, Strobino D, Bartlett L. Development and Validation of an Index to Measure the Quality of Facility-Based Labor and Delivery Care Processes in Sub-Saharan Africa. *PLoS One*. 2015;10(6):e0129491.

Appendix 2. Validation Analyses

Appendix 2a. Construct Validity Analysis: Tetrachoric Correlation Matrices of the 7-item Interpersonal Abuse Scale for Nigeria, Ghana, and Guinea

Tetrachoric correlation matrix:	Shouted	Insulted	Scolded	Mocked	Threatened	Negative comments	Physical abuse
NIGERIA (N=407)							
Shouted	1.00						
Insulted	0.44	1.00					
Scolded	0.59	0.62	1.00				
Mocked	0.43	0.61	0.41	1.00			
Threatened	0.47	0.39	0.36	0.54	1.00		
Negative comments	0.53	0.55	0.52	0.32	0.31	1.00	
Physical abuse	0.60	0.44	0.33	0.42	0.22	0.32	1.00
GHANA (N=912)							
Shouted	1.00						
Insulted	0.58	1.00					
Scolded	0.44	0.24	1.00				
Mocked	0.57	0.60	0.29	1.00			
Threatened	0.32	0.08	0.30	0.45	1.00		
Negative comments	0.57	0.49	0.34	0.43	0.46	1.00	
Physical abuse	0.58	0.58	0.33	0.29	0.30	0.37	1.00
GUINEA (N=655)							
Shouted	1.00						
Insulted	0.51	1.00					
Scolded	0.43	0.40	1.00				
Mocked	0.45	0.44	0.41	1.00			
Threatened	0.31	0.29	0.39	0.38	1.00		
Negative comments	0.23	0.37	0.24	0.57	0.33	1.00	
Physical abuse	0.48	0.54	0.26	0.39	0.18	0.07	1.00

Appendix 2b. Construct Validity Analysis: Spearman-Rank Correlations Between Mistreatment Measures by Country

Measurement	Interpersonal Abuse Scale	Exams & Procedures Index	Unsupportive Birth Environment Index
Nigeria (N=407)			
Interpersonal Abuse Scale	1.00		
Exams & Procedures Index	0.19	1.00	
Unsupportive Birth Environment Index	-0.06	-0.06	1.00
Ghana (N=912)			
Interpersonal Abuse Scale	1.00		
Exams & Procedures Index	0.13	1.00	
Unsupportive Birth Environment Index	0.002	0.10	1.00
Guinea (N=655)			
Interpersonal Abuse Scale	1.00		
Exams & Procedures Index	0.06	1.00	
Unsupportive Birth Environment Index	0.03	-0.08	1.00

*p<0.05

Appendix 2c. Criterion Validity Analysis: Bivariate Logistic Regressions of Measures of Satisfaction^a on mistreatment measure score by measure and country

	Interpersonal Abuse Scale OR (95% CI)			Exams & Procedures Index OR (95% CI)			Unsupportive Birth Environment Index OR (95% CI)		
	Nigeria (n=331)	Ghana (n=771)	Guinea (n=423)	Nigeria (n=331)	Ghana (n=771)	Guinea (n=423)	Nigeria (n=331)	Ghana (n=771)	Guinea (n=423)
Overall satisfaction with care									
Satisfied/neutral	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
Dissatisfied	0.99 (0.81-1.21)	1.54* (1.20-1.99)	1.42* (1.08-1.85)	1.00 (0.85-1.18)	0.96 (0.52-1.75)	0.73 (0.33-1.60)	0.97 (0.62-1.54)	0.93 (0.68-1.27)	1.21 (0.88-1.68)
Would deliver in same facility in future									
No	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
Yes	0.90 (0.56-1.44)	0.89 (0.74-1.06)	0.76* (0.60-0.96)	0.72 (0.32-1.58)	1.08 (0.79-1.47)	1.09 (0.63-1.86)	1.47 (0.98-2.23)	1.01 (0.85-1.21)	0.81 (0.2-1.05)

^a Based on women's reports from subset of linked labor observation and community survey data (Nigeria: N=331; Ghana: N=771; Guinea: N=423). Analyses adjusted for clustering within facilities.

* p<0.05

Appendix 3. Reliability/Internal Consistency: Distribution of Co-Occurring Mistreatment Items for Each Item by Measure

Appendix 3a. Distribution of Co-Occurring Mistreatment Items for Each Item by Measure, Nigeria (N=407, unless otherwise noted)*

Mistreatment Item	Number of Co-Occurring Items (%) ^a											Total (n)	n (%) without item ^b	
	0	1	2	3	4	5	6	7	8	9	10			11
<i>Interpersonal Abuse Scale (7 items)</i>														
Shouted	20.1	28.5	21.0	12.6	10.3	5.6	1.9	--	--	--	--	--	214	193 (47.4)
Insulted	7.3	13.4	14.6	20.7	24.4	14.6	4.9	--	--	--	--	--	82	325 (79.9)
Scolded	7.4	19.4	24.1	16.7	18.5	10.2	3.7	--	--	--	--	--	108	299 (73.5)
Mocked	3.5	20.7	25.3	16.1	17.2	12.6	4.6	--	--	--	--	--	87	320 (78.6)
Negative comments	2.1	17.0	25.5	17.0	17.0	12.8	8.5	--	--	--	--	--	47	360 (88.5)
Threatened	4.5	14.9	23.9	19.4	19.4	11.9	6.0	--	--	--	--	--	67	340 (83.5)
Physical abuse	12.7	22.7	20.9	13.6	15.5	10.9	3.6	--	--	--	--	--	110	297 (73.0)
Total Score (%) women with score of ≥ 1	--	28.2	28.2	19.4	9.9	8.4	4.4	1.5	--	--	--	--	273	134 (32.9)
<i>Exams & Procedures Index (3 items) (N=350)</i>														
Informed consent	18.2	32.4	49.5	--	--	--	--	--	--	--	--	--	275	75 (21.4)
Exposed	7.2	41.5	51.3	--	--	--	--	--	--	--	--	--	265	85 (24.5)
Confidential information	0.60	17.5	81.9	--	--	--	--	--	--	--	--	--	166	184 (52.3)
Total Score (%) women with score of ≥ 1	--	20.0	32.6	38.9	--	--	--	--	--	--	--	--	320	30 (8.6)
<i>Unsupportive Birth Environment Index (12 items)</i>														
Pain relief	--	0.3	9.3	47.9	38.1	3.7	0.8	--	--	--	--	--	378	29 (7.1)
No interpreter	--	--	--	--	33.3	33.3	33.3	--	--	--	--	--	3	404 (99.3)
No staff present at birth	--	--	9.7	29.0	22.6	32.3	6.5	--	--	--	--	--	31	376 (92.4)
Not offered labor companion	--	0.8	11.0	46.8	37.3	3.3	0.8	--	--	--	--	--	391	16 (3.9)
Neglect	--	--	--	--	50.0	33.3	16.7	--	--	--	--	--	6	401 (98.5)
Bribe	--	--	--	21.4	57.1	14.3	7.1	--	--	--	--	--	14	393 (96.6)
Clean up blood/fluids	--	--	--	16.7	33.3	33.3	16.7	--	--	--	--	--	6	401 (98.5)
Fluids	--	--	0.6	17.7	74.9	6.3	0.6	--	--	--	--	--	175	232 (57.0)
Mobilize	--	1.1	8.1	48.3	38.1	3.7	0.8	--	--	--	--	--	381	26 (6.4)
No curtains/partitions	--	1.1	8.3	46.6	39.4	3.6	0.8	--	--	--	--	--	360	47 (11.6)

Mistreatment Item	Number of Co-Occurring Items (%) ^a											Total (n)	n (%) without item ^b	
	0	1	2	3	4	5	6	7	8	9	10			11
No bed	--	--	20.0	--	--	20.0	40.0	20.0	--	--	--	--	5	402 (98.8)
Shared bed	--	--	--	--	--	100.0	--	--	--	--	--	--	1	406 (99.8)
Total Score (%) women with score of ≥ 1	--	--	1.5	11.8	46.7	35.8	3.4	0.7	--	--	--	--	407	n (%) scoring 0 0 (0.0)

Appendix 3a

* Table shows the percent distributions of co-occurring mistreatment items to evaluate internal consistency of items within each measure. Internally consistent measures have higher proportions of items that occur with at least one or several other items in a measure.

^a Proportions sum by row; ^b Proportions sum by column

Appendix 3b. Distribution of Co-Occurring Mistreatment Items for Each Item by Measure, Ghana (N=912, unless otherwise noted)*

Mistreatment Item	Number of Co-Occurring Items (%) ^a											Total (n)	n (%) without item ^b	
	0	1	2	3	4	5	6	7	8	9	10			11
<i>Interpersonal Abuse Scale (7 items)</i>														
Shouted	42.9	33.3	15.9	5.3	2.1	0.5	--	--	--	--	--	--	189	723 (79.3)
Insulted	15.4	11.5	42.3	23.1	3.9	3.9	--	--	--	--	--	--	26	886 (97.2)
Scolded	31.1	36.5	23.0	6.8	2.7	--	--	--	--	--	--	--	74	838 (91.9)
Mocked	9.1	31.8	31.8	9.1	13.6	4.6	--	--	--	--	--	--	22	890 (97.6)
Negative comments	32.7	36.7	14.3	8.2	6.1	2.0	--	--	--	--	--	--	49	863 (94.6)
Threatened	16.7	35.2	29.6	11.1	5.6	1.9	--	--	--	--	--	--	54	858 (94.1)
Physical abuse	27.0	33.8	23.0	9.5	5.4	1.4	--	--	--	--	--	--	74	838 (91.9)
Total Score (%) women with score of ≥ 1	--	54.2	28.5	12.0	3.5	1.4	0.4	--	--	--	--	--	284	n (%) scoring 0 628 (68.9)
<i>Exams & Procedures Index (3 items) (n=797)</i>														
Informed consent	90.8	6.2	2.9	--	--	--	--	--	--	--	--	--	578	219 (27.5)
Exposed	11.3	56.6	32.1	--	--	--	--	--	--	--	--	--	53	744 (93.4)
Confidential information	18.2	30.3	51.5	--	--	--	--	--	--	--	--	--	33	764 (95.9)
Total Score (%) women with score of ≥ 1	--	90.7	6.4	2.9	--	--	--	--	--	--	--	--	592	n (%) scoring 0 205 (25.7)
<i>Unsupportive Birth Environment Index (12 items)</i>														
Pain relief	0.4	15.9	35.4	39.1	8.7	0.5	--	--	--	--	--	--	785	127 (13.9)
No interpreter	--	--	25.0	33.3	33.3	8.3	--	--	--	--	--	--	12	900 (98.7)
No staff present at birth	--	--	17.7	11.8	58.8	11.8	--	--	--	--	--	--	17	895 (98.1)
Not offered labor companion	4.9	17.9	32.7	36.1	8.0	0.5	--	--	--	--	--	--	854	58 (6.4)
Neglect	--	--	50.0	50.0	--	--	--	--	--	--	--	--	2	910 (99.8)
Bribe	--	--	50.0	16.7	33.3	--	--	--	--	--	--	--	6	906 (99.3)
Clean up blood/fluids	--	--	--	--	--	--	--	--	--	--	--	--	0	912 (100)
Fluids	--	3.0	15.8	63.3	16.8	1.0	--	--	--	--	--	--	398	514 (56.4)
Mobilize	0.4	7.8	37.6	43.9	9.7	0.6	--	--	--	--	--	--	692	220 (24.1)
No curtains/partitions	--	5.1	15.9	43.5	34.1	1.5	--	--	--	--	--	--	138	774 (95.9)
No bed	--	12.5	25.0	50.0	12.5	--	--	--	--	--	--	--	8	904 (99.1)
Shared bed	--	--	--	63.6	18.2	18.2	--	--	--	--	--	--	11	901 (98.8)
														n (%) scoring 0

Mistreatment Item	Number of Co-Occurring Items (%) ^a											Total (n)	n (%) without item ^b	
	0	1	2	3	4	5	6	7	8	9	10			11
Total Score (%) women with score of ≥ 1	-	5.3	19.3	33.4	34.1	7.5	0.4	--	--	--	--	--	912	0 (0.0)

Appendix 3b

* Table shows the percent distributions of co-occurring mistreatment items to evaluate internal consistency of items within each measure. Internally consistent measures have higher proportions of items that occur with at least one or several other items in a measure.

^a Proportions sum by row; ^b Proportions sum by column

Appendix 3c. Distribution of Co-Occurring Mistreatment Items for Each Item by Measure, Guinea (N=655, unless otherwise noted)*

Mistreatment Item	Number of Co-Occurring Items (%) ^a											Total (n)	n (%) without item ^b	
	0	1	2	3	4	5	6	7	8	9	10			11
<i>Interpersonal Abuse Scale (7 items)</i>														
Shouted	40.3	34.5	14.4	7.2	2.9	--	0.7	--	--	--	--	--	139	516 (78.8)
Insulted	11.1	27.8	11.1	38.9	5.6	--	5.6	--	--	--	--	--	18	637 (97.3)
Scolded	38.5	30.8	11.5	12.8	5.1	--	1.3	--	--	--	--	--	78	577 (88.1)
Mocked	25.0	25.0	28.9	11.5	7.7	--	1.9	--	--	--	--	--	52	603 (92.1)
Negative comments	26.7	40.0	6.7	13.3	6.7	--	6.7	--	--	--	--	--	15	640 (97.7)
Threatened	40.9	18.2	9.1	18.2	9.1	--	4.6	--	--	--	--	--	22	633 (96.6)
Physical abuse	40.4	32.7	13.5	8.7	3.9	--	1.0	--	--	--	--	--	104	551 (84.1)
Total Score (%) women with score of ≥ 1	--	60.0	26.5	6.9	4.6	1.5	--	0.4	--	--	--	--	260	395 (60.3)
<i>Exams & Procedures Index (3 items) (n=391)</i>														
Informed consent	63.3	33.2	3.6	--	--	--	--	--	--	--	--	--	196	195 (49.9)
Exposed	32.7	61.1	6.2	--	--	--	--	--	--	--	--	--	113	278 (71.1)
Confidential information	55.3	26.3	18.4	--	--	--	--	--	--	--	--	--	38	353 (90.3)
Total Score (%) women with score of ≥ 1	--	69.7	27.6	2.7	--	--	--	--	--	--	--	--	261	130 (33.3)
<i>Unsupportive Birth Environment Index (11 items)</i>														
Pain relief	0.5	17.5	38.4	25.7	14.3	3.3	0.4	--	--	--	--	--	573	82 (12.5)
No interpreter	--	--	--	--	50.0	50.0	--	--	--	--	--	--	2	653 (99.7)

Mistreatment Item	Number of Co-Occurring Items (%) ^a											Total (n)	n (%) without item ^b	
	0	1	2	3	4	5	6	7	8	9	10			11
No staff present at birth	--	5.7	22.6	30.2	34.0	7.6	--	--	--	--	--	--	53	602 (91.9)
Not offered labor companion	2.3	18.0	37.7	24.4	13.6	3.3	0.4	--	--	--	--	--	579	76 (11.6)
Neglect	--	--	--	--	60.0	40.0	--	--	--	--	--	--	5	650 (99.2)
Bribe	--	--	9.5	23.8	52.4	11.9	2.4	--	--	--	--	--	42	613 (93.6)
Clean up blood/fluids	--	--	--	--	--	100.0	--	--	--	--	--	--	1	654 (99.9)
Fluids	0.5	4.0	23.8	34.7	27.2	8.9	1.0	--	--	--	--	--	202	453 (69.2)
Mobilize	--	1.6	16.8	40.0	31.2	8.8	1.6	--	--	--	--	--	125	530 (80.9)
No curtains/partitions	0.5	11.5	36.1	29.1	18.5	3.9	0.5	--	--	--	--	--	416	239 (36.5)
No bed	--	3.3	3.3	26.7	23.3	33.3	10.0	--	--	--	--	--	30	625 (95.4)
Shared bed	--	2.1	21.0	18.2	38.5	18.9	1.4	--	--	--	--	--	143	512 (78.2)
Total Score (%) women with score of ≥ 1	--	3.4	20.8	36.3	23.7	12.7	2.9	0.3	--	--	--	--	655	0 (0.0)

Appendix 3c

* Table shows the percent distributions of co-occurring mistreatment items to evaluate internal consistency of items within each measure. Internally consistent measures have higher proportions of items that occur with at least one or several other items in a measure.

^a Proportions sum by row; ^b Proportions sum by column

Appendix 4. Parallel Analysis Figures for 7-Item Interpersonal Abuse Scale from Principal Components Analysis in Nigeria, Ghana, and Guinea

