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

Blair O Berger, Donna M Strobino, Hedieh Mehrtash, Meghan A Bohren, Kwame Adu-Bonsaffoh, Hannah H Leslie, Theresa Azonima Irinyenikan, Thae Maung Maung, Mamadou Diouf Balde, Özge Tunçalp

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) 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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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’).\textsuperscript{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.\textsuperscript{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’\textsuperscript{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.\textsuperscript{15}

Recent prevalence estimates of mistreatment range widely across studies from 13% to 98%.\textsuperscript{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.\textsuperscript{17 19 36–35} 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.\textsuperscript{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.\textsuperscript{22 41–45} 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.\textsuperscript{22 41–45}

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.\textsuperscript{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.\textsuperscript{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).

\textbf{METHODS}

\textbf{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.\textsuperscript{45–44}

\textbf{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

\textsuperscript{8}
survey tool. The second study phase implemented and validated the instruments. Details of the WHO study methodology are described elsewhere.\textsuperscript{10} \textsuperscript{47} \textsuperscript{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 in the event of repeated occurrences of the same type of mistreatment.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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-Olkin 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.
Interpersonal Abuse Scale development: psychometric analysis

The Interpersonal Abuse Scale was developed using a psychometric approach outlined by DeVellis and Netemeyer et al. Exploratory factor analysis (EFA) was conducted separately by country to assess the construct validity and reliability of the scale. 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. 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. EFA was then performed using iterated principal factor estimation. To establish unidimensionality, items with either low (λ <0.40) or very high (λ >0.90, indicating possible redundancy) factor loadings were omitted.

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. 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. 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.

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, the indexing approach outlined in the Consensus-based Standards for selection of health Measurement Instruments initiative (COSMIN), and existing indices measuring quality of care in maternal and reproductive health. 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.

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<0.05) as evidence of significant correlation. 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. 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 I 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.

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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
</tr>
<tr>
<td><strong>Individual characteristics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years): mean (SD)</td>
<td>29.3 (5.4)</td>
<td>28.4 (5.9)</td>
<td>23.8 (5.6)</td>
</tr>
<tr>
<td><strong>Education level</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None/primary</td>
<td>22 (5.4)</td>
<td>293 (32.1)</td>
<td>438 (66.9)</td>
</tr>
<tr>
<td>Secondary</td>
<td>176 (43.2)</td>
<td>462 (50.7)</td>
<td>152 (23.2)</td>
</tr>
<tr>
<td>Tertiary</td>
<td>204 (50.1)</td>
<td>134 (14.7)</td>
<td>43 (6.6)</td>
</tr>
<tr>
<td>Vocational or other</td>
<td>5 (1.2)</td>
<td>23 (2.5)</td>
<td>22 (3.4)</td>
</tr>
<tr>
<td><strong>Marital status</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>386 (94.8)</td>
<td>735 (80.6)</td>
<td>612 (93.4)</td>
</tr>
<tr>
<td>Not married</td>
<td>21 (5.2)</td>
<td>177 (19.4)</td>
<td>43 (6.6)</td>
</tr>
<tr>
<td><strong>Number of previous births</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>163 (40.1)</td>
<td>343 (37.6)</td>
<td>235 (35.9)</td>
</tr>
<tr>
<td>1</td>
<td>110 (27.0)</td>
<td>249 (27.3)</td>
<td>143 (21.8)</td>
</tr>
<tr>
<td>2+</td>
<td>134 (32.9)</td>
<td>320 (35.1)</td>
<td>277 (42.3)</td>
</tr>
<tr>
<td><strong>Provider characteristics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary labour attendant *</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nurse/midwife</td>
<td>360 (88.5)</td>
<td>458 (69.9)</td>
<td>762 (83.6)</td>
</tr>
<tr>
<td>Doctor</td>
<td>20 (4.9)</td>
<td>103 (15.7)</td>
<td>101 (11.1)</td>
</tr>
<tr>
<td>Trainee</td>
<td>22 (5.4)</td>
<td>75 (11.5)</td>
<td>37 (4.1)</td>
</tr>
<tr>
<td>None/other</td>
<td>5 (1.2)</td>
<td>19 (2.9)</td>
<td>12 (1.3)</td>
</tr>
<tr>
<td>Staff member present at birth**†</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nurse/midwife</td>
<td>306 (81.4)</td>
<td>667 (74.5)</td>
<td>414 (68.8)</td>
</tr>
<tr>
<td>Doctor</td>
<td>66 (17.6)</td>
<td>216 (24.1)</td>
<td>156 (25.9)</td>
</tr>
<tr>
<td>Trainee</td>
<td>1 (0.3)</td>
<td>11 (1.2)</td>
<td>32 (5.3)</td>
</tr>
<tr>
<td>Other</td>
<td>3 (0.8)</td>
<td>1 (0.1)</td>
<td>--</td>
</tr>
<tr>
<td><strong>Characteristics of childbirth</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mode of birth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vaginal</td>
<td>380 (93.4)</td>
<td>765 (83.4)</td>
<td>563 (86.0)</td>
</tr>
<tr>
<td>Caesarean</td>
<td>27 (6.6)</td>
<td>147 (16.1)</td>
<td>92 (14.1)</td>
</tr>
<tr>
<td>Companion presence during labour and/or childbirth</td>
<td>78 (19.2)</td>
<td>102 (11.2)</td>
<td>66 (10.1)</td>
</tr>
<tr>
<td>Type of companion present at birth (n=163)</td>
<td>n=51</td>
<td>n=82</td>
<td>n=31</td>
</tr>
<tr>
<td>Husband/partner</td>
<td>21 (41.2)</td>
<td>62 (75.6)</td>
<td>--</td>
</tr>
<tr>
<td>Family/friend</td>
<td>26 (51.0)</td>
<td>20 (24.4)</td>
<td>30 (97.0)</td>
</tr>
<tr>
<td>Other</td>
<td>4 (7.8)</td>
<td>--</td>
<td>1 (3.2)</td>
</tr>
<tr>
<td>Time of childbirth‡</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day</td>
<td>201 (49.4)</td>
<td>390 (42.8)</td>
<td>280 (42.8)</td>
</tr>
<tr>
<td>Night</td>
<td>206 (50.6)</td>
<td>522 (57.2)</td>
<td>375 (57.3)</td>
</tr>
<tr>
<td>Day of childbirth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weekday</td>
<td>309 (75.9)</td>
<td>673 (73.8)</td>
<td>513 (78.3)</td>
</tr>
<tr>
<td>Weekend</td>
<td>98 (24.1)</td>
<td>239 (26.2)</td>
<td>142 (21.7)</td>
</tr>
</tbody>
</table>

*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 Guinea); 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 tool.47
<table>
<thead>
<tr>
<th>Mistreatment Dimension</th>
<th>Item</th>
<th>WHO mistreatment labour observation tool question</th>
<th>Item scoring coding*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interpersonal Abuse Scale (7 items)</td>
<td>1. Shouted/screamed at</td>
<td>Was the woman shouted or screamed at?</td>
<td>0: Q=1 (no), 9 (DK)</td>
</tr>
<tr>
<td></td>
<td>2. Insulted</td>
<td>Was the woman insulted?</td>
<td>0: Q=1 (no), 9 (DK)</td>
</tr>
<tr>
<td></td>
<td>3. Scolded</td>
<td>Was the woman scolded?</td>
<td>0: Q=1 (no), 9 (DK)</td>
</tr>
<tr>
<td></td>
<td>4. Mocked</td>
<td>Was the woman mocked?</td>
<td>0: Q=1 (no), 9 (DK)</td>
</tr>
<tr>
<td></td>
<td>5. Negative comments</td>
<td>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)?</td>
<td>0: Q=1 (no), 9 (DK)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1: Q=2 (yes) on at least one negative comment</td>
</tr>
<tr>
<td></td>
<td>6. Threatened</td>
<td>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)?</td>
<td>0: Q=1 (no), 9 (DK)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1: Q=2 (yes) on at least one threat</td>
</tr>
<tr>
<td></td>
<td>7. Physical abuse</td>
<td>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</td>
<td>0: Q=1 (no), 9 (DK)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1: Q=2 (yes) on at least one physical abuse item</td>
</tr>
<tr>
<td>Deleted Items</td>
<td>Hissed</td>
<td>Was the woman hissed at?</td>
<td>Reason for deletion</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Nigerian-specific item</td>
</tr>
<tr>
<td></td>
<td>Blamed</td>
<td>Was the woman blamed for her or her baby’s poor outcome?</td>
<td>Low frequency, low factor loading</td>
</tr>
<tr>
<td>Exams &amp; Procedures Index (3 items)</td>
<td>1. Informed consent†</td>
<td>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</td>
<td>0: Q=2 (yes), 9 (DK)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1: Q=1 (no) on either first vaginal exam or procedure</td>
</tr>
<tr>
<td></td>
<td>2. Exposed‡</td>
<td>Did a staff member conduct a vaginal exam in a way that others could see her genitals or breasts?</td>
<td>0: Q=1 (no), 9 (DK)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1: Q=2 (yes)</td>
</tr>
<tr>
<td></td>
<td>3. Confidential information‡</td>
<td>(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)?</td>
<td>0: Q=1 (no), 9 (DK)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1: Q=2 (yes)</td>
</tr>
<tr>
<td>Unsupportive Birth Environment Index (12 items)</td>
<td>Support</td>
<td>1. Pain relief</td>
<td>Was the woman offered any form of pain relief or was the woman given pain relief if she requested it?</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1: Q=1 (no)</td>
</tr>
<tr>
<td></td>
<td>2. No interpreter</td>
<td>Was an interpreter used (if the woman’s primary birth attendant did not speak same language as the woman?)</td>
<td>0: Q=2 (yes), 9 (DK), 8(NA)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1: Q=1 (no)</td>
</tr>
<tr>
<td></td>
<td>3. No staff present at birth</td>
<td>Was a staff member present when the baby came out?</td>
<td>0: Q=2 (yes), 9 (DK)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1: Q=1 (no)</td>
</tr>
<tr>
<td></td>
<td>4. Not offered labour companion</td>
<td>Was the woman offered to have a labour companion during labour and delivery?</td>
<td>0: Q=2 (yes), 9 (DK)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1: Q=1 (no)</td>
</tr>
<tr>
<td></td>
<td>5. Neglect</td>
<td>Did the woman request medical attention from a health worker that was not responded to?</td>
<td>0: Q=1 (no), 9 (DK)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1: Q=2 (yes)</td>
</tr>
<tr>
<td></td>
<td>Birth environment</td>
<td>6. Bribe</td>
<td>At any time, did staff suggest or ask the woman (or companion) for a bribe, informal payment or gift?</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1: Q=2 (yes)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7. Clean up blood/ fluids</td>
<td>At any time, was the woman instructed to clean up blood, urine, faeces or amniotic fluid?</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1: Q=2 (yes)</td>
</tr>
</tbody>
</table>

Continued
developed based on five domains of the WHO typology on mistreatment of women during childbirth. 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 based on a specific set of data. 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 for frequency in all three countries.

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.

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’
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 abdomen, other use of physical force.

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

<table>
<thead>
<tr>
<th>Item</th>
<th>Nigeria (n=407) %</th>
<th>Ghana (n=912) %</th>
<th>Guinea (n=655) %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interpersonal Abuse Scale (7 items)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Shouted/screamed at</td>
<td>52.6</td>
<td>20.7</td>
<td>21.2</td>
</tr>
<tr>
<td>2. Insulted</td>
<td>20.2</td>
<td>2.9</td>
<td>2.8</td>
</tr>
<tr>
<td>3. Scolded</td>
<td>26.5</td>
<td>8.1</td>
<td>11.9</td>
</tr>
<tr>
<td>4. Mocked</td>
<td>21.4</td>
<td>2.4</td>
<td>7.9</td>
</tr>
<tr>
<td>5. Negative comments *</td>
<td>11.6</td>
<td>5.4</td>
<td>2.3</td>
</tr>
<tr>
<td>6. Threatened †</td>
<td>16.5</td>
<td>5.9</td>
<td>3.4</td>
</tr>
<tr>
<td>7. Physical abuse ‡</td>
<td>27.0</td>
<td>8.1</td>
<td>15.9</td>
</tr>
</tbody>
</table>

| Mean Scale Score (±SD)                    |                   |                 |                 |
| Maximum score: 7 points                   | 1.8 (1.8) (range: 0–7) | 0.5 (1.0) (range: 0–6) | 0.7 (1.0) (range: 0–7) |
| % of women>country mean                   | 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             |

| Mean Index Score (±SD)                    |                   |                 |                 |
| Maximum score: 3 points                   | 2.0 (0.97) (range: 0–3) | 0.8 (0.60) (range: 0–3) | 0.9 (0.76) (range: 0–3) |
| % of women>country mean                   | 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            |

| Mean Index Score (±SD)                    |                   |                 |                 |
| Maximum score: 12 points                  | 4.3 (0.81) (range: 2–7) | 3.2 (1.0) (range: 1–6) | 3.3 (1.1) (range: 1–7) |
| % of women>country mean                   | 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.
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. 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. 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. 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>
<thead>
<tr>
<th>Item (7 items)</th>
<th>Nigeria (N=407)</th>
<th>Ghana (N=912)</th>
<th>Guinea (N=655)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Factor loading</td>
<td>Uniqueness</td>
<td>Factor loading</td>
</tr>
<tr>
<td>1. Shouted/screamed at</td>
<td>0.77</td>
<td>0.41</td>
<td>0.84</td>
</tr>
<tr>
<td>2. Insulted</td>
<td>0.77</td>
<td>0.40</td>
<td>0.71</td>
</tr>
<tr>
<td>3. Scolded</td>
<td>0.72</td>
<td>0.48</td>
<td>0.48</td>
</tr>
<tr>
<td>4. Mocked</td>
<td>0.67</td>
<td>0.55</td>
<td>0.68</td>
</tr>
<tr>
<td>5. Negative comments</td>
<td>0.56</td>
<td>0.68</td>
<td>0.46</td>
</tr>
<tr>
<td>6. Threatened</td>
<td>0.64</td>
<td>0.60</td>
<td>0.69</td>
</tr>
<tr>
<td>7. Physical abuse *</td>
<td>0.58</td>
<td>0.67</td>
<td>0.64</td>
</tr>
</tbody>
</table>

Internal reliability

<table>
<thead>
<tr>
<th></th>
<th>Internal reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>KR-20 coefficient</td>
<td>0.71</td>
</tr>
<tr>
<td>Deleted items</td>
<td></td>
</tr>
<tr>
<td>1. Was the woman blamed for her or her baby’s poor health outcomes?</td>
<td>Omitted due to low frequency</td>
</tr>
<tr>
<td>2. Was the woman hissed at?</td>
<td>Nigeria-specific item</td>
</tr>
</tbody>
</table>

*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.
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 care and person-centred maternity care. However, the extent to which women's reports of satisfaction are consistently related to mistreatment, particularly third-party labour observations

<table>
<thead>
<tr>
<th>Table 5</th>
<th>Percent distribution of women experiencing a mistreatment item by mistreatment score and country*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mistreatment item</td>
<td>Nigeria (N=407)</td>
</tr>
<tr>
<td></td>
<td>Low score</td>
</tr>
<tr>
<td>Interpersonal Abuse Scale (7 items)</td>
<td></td>
</tr>
<tr>
<td>Shouted</td>
<td>20.4</td>
</tr>
<tr>
<td>Insulted</td>
<td>2.8</td>
</tr>
<tr>
<td>Scolded</td>
<td>3.8</td>
</tr>
<tr>
<td>Mocked</td>
<td>1.4</td>
</tr>
<tr>
<td>Negative comments</td>
<td>0.5</td>
</tr>
<tr>
<td>Threatened</td>
<td>1.4</td>
</tr>
<tr>
<td>Physical abuse</td>
<td>6.6</td>
</tr>
<tr>
<td>Total women (n)</td>
<td>211</td>
</tr>
<tr>
<td>Country mean score (unweighted) (SD)</td>
<td>1.8 (1.8)</td>
</tr>
<tr>
<td>(range: 0–7)</td>
<td></td>
</tr>
<tr>
<td>% above country mean ('high')</td>
<td>48.2</td>
</tr>
<tr>
<td>Exams &amp; Procedures Index (3 items)</td>
<td>(n=350)</td>
</tr>
<tr>
<td>Informed consent</td>
<td>65.0</td>
</tr>
<tr>
<td>Exposed</td>
<td>60.3</td>
</tr>
<tr>
<td>Confidential information</td>
<td>14.0</td>
</tr>
<tr>
<td>Total Women (n)</td>
<td>214</td>
</tr>
<tr>
<td>Country mean score (unweighted) (SD)</td>
<td>2.0 (0.97)</td>
</tr>
<tr>
<td>(range: 0–3)</td>
<td></td>
</tr>
<tr>
<td>% above country mean ('high')</td>
<td>38.9</td>
</tr>
<tr>
<td>Unsupportive Birth Environment Index (12 items)</td>
<td></td>
</tr>
<tr>
<td>Pain relief</td>
<td>88.9</td>
</tr>
<tr>
<td>No interpreter</td>
<td>0.0</td>
</tr>
<tr>
<td>No staff present at birth</td>
<td>4.9</td>
</tr>
<tr>
<td>Not offered labour companion</td>
<td>93.9</td>
</tr>
<tr>
<td>Neglect</td>
<td>0.0</td>
</tr>
<tr>
<td>Bribe</td>
<td>1.2</td>
</tr>
<tr>
<td>Clean up blood/fluids</td>
<td>0.4</td>
</tr>
<tr>
<td>Fluids</td>
<td>13.1</td>
</tr>
<tr>
<td>Mobilise</td>
<td>89.8</td>
</tr>
<tr>
<td>No curtains/partitions</td>
<td>82.8</td>
</tr>
<tr>
<td>No bed</td>
<td>0.4</td>
</tr>
<tr>
<td>Shared bed</td>
<td>0.0</td>
</tr>
<tr>
<td>Total women (n)</td>
<td>244</td>
</tr>
<tr>
<td>Country mean score (unweighted) (SD)</td>
<td>4.3 (0.81)</td>
</tr>
<tr>
<td>(range: 2–7)</td>
<td></td>
</tr>
<tr>
<td>% above country mean ('high')</td>
<td>40.1</td>
</tr>
</tbody>
</table>

*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.
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.9 84 Taken together, our satisfaction as an outcome measure— informed by expecta-
tions 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 generalis-
ability. 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 valida-
tion 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 like-
lihood of mistreatment by various provider character-
istics.11 29 86 Assessing items in the Interpersonal Abuse Scale as binary rather than incorporating repeated incid-
ents could have reduced data variability. Using continu-
ous 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 woman’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 method-
ological 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 instru-
ment, 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 interven-
tions that impact both interpersonal aspects of care and more systemic, structural aspects of the process of care. The measures can also translate evidence-based recom-
mandations 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
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 Tha Maung Maung http://orcid.org/0000-0002-1265-3813

REFERENCES
13 Rominski SD, Loni 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.


51 Bohren MA, Vogel JP, Tunçalo 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;2:420–55.


53 Staats Corp LP. Stata/MP 13.0. College Station, TX, 2013.


