

Developing and Validating a Perinatal Depression Screening Tool in Bungoma County,

Kenya

by

Hawa Tuli

Duke Global Health Institute
Duke University

Date: _____

Approved:

Eric Green, Supervisor

Brandon Kohrt

Lavanya Vasudevan

Thesis submitted in partial fulfillment of
the requirements for the degree of
Master of Science in the Duke Global Health Institute
in the Graduate School of Duke University

2016

ABSTRACT

Developing and Validating a Perinatal Depression Screening Tool in Bungoma County,

Kenya

by

Hawa Tuli

Duke Global Health Institute
Duke University

Date: _____

Approved:

Eric Green, Supervisor

Brandon Kohrt

Lavanya Vasudevan

An abstract of a thesis submitted in partial
fulfillment of the requirements for the degree
of Master of Science in the Duke Global Health Institute
in the Graduate School of Duke University

2016

Copyright by
Hawa Tuli
2016

Abstract

Background: Depression-screening tools exist and are widely used in Western settings. There have been few studies done to explore whether or not existing tools are valid and effective for use in sub-Saharan Africa. There is, thus, a critical need to develop and validate perinatal depression screening tools for use in different settings by personnel of varying professional levels. Our study aimed to develop and validate a perinatal depression-screening tool in rural Kenya.

Methods: We conducted free listing and card sorting exercises with a purposive sample of 12 women and 38 community health volunteers (CHVs) living in a rural community to explore the manifestations of perinatal depression in that setting. We used the information obtained to produce a locally relevant depression-screening tool that comprised of existing Western psychiatric concepts and locally derived items. Subsequently, we administered the novel depression-screening tool and two existing screening tools (the Edinburgh Postnatal Depression Scale and the Patient Health Questionnaire-9) to a random sample of 193 women and compared the results of the screening tools with that of a gold standard structured clinical interview to test diagnostic accuracy.

Results: The free listing and card sorting exercise produced a set of 60 screening items. Of the items in this set, we identified the 9 items that most accurately classified cases

and non-cases. This 9-item scale had a sensitivity/specificity of 0.90/0.90. This compared to 0.70/0.72 and 0.70/0.73 for the EPDS and PHQ-9, respectively. Across all three tools, internal consistency reliability ranged from 0.77 to 0.81 and test-retest reliability ranged from 0.57 to 0.67. The prevalence of depression ranged from 5.2 to 6.2 percent depending on the clinical reference standard used.

Conclusions: While the EPDS and PHQ-9 are valid and reliable screening tools for perinatal depression in rural Western Kenya, a new 9-item locally anchored tool called the Perinatal Depression Screening (PDEPS) that incorporates local idioms of distress with existing Western psychiatric concepts may be a better alternative. We found the prevalence of depression in this region to be lower than other published estimates for African and other low-income countries.

Contents

Abstract.....	iv
List of Tables	viii
Acknowledgements	ix
Introduction	1
Methods.....	3
Setting and Participants.....	3
Measures	5
Screening Survey	5
Edinburgh Postnatal Depression Scale	6
Patient Health Questionnaire-9.....	7
Criterion Reference: Structured Clinical Interview for DSM-5	7
Alternate Criterion Reference: Local Diagnosis	8
Procedures	9
Study strengths and limitations	33
Conclusion	35
Appendix A.....	36
Appendix B	37
Appendix C.....	38
Appendix D.....	39
Appendix E	40

Appendix F	41
Appendix G.....	42
Appendix H	43
Appendix I	44
References	45

List of Tables

Table 1: Participant demographic information.....	23
Table 2: Diagnostic results by maternal status	24
Table 3: Items in Perinatal Depression Screening (PDEPS)	26
Table 4: Diagnostic validity	27
Table 5: Mean scale scores for cases and non-cases to evaluate convergent validity of PDEPS and existing scales	28
Table 6: Mean scale scores for cases and non-cases to evaluate discriminant validity of PDEPS and existing scales	28
Table 7: Reliability of PDEPS and existing scales.....	29

Acknowledgements

I would first and foremost like to thank my research mentor, Dr. Eric Green, whose unwavering support and guidance made this project possible. I would also like to thank Christina Schmidt, Carlyne Khwituta, Sarah Wanyonyi, Katherine King, Edna Sang, Teresa Kangogo, Mary Ogina, Medina Nyongesa, Eileen Mukonye, Alice Seurey, Maximilla Chivinh, Everline Walutila, and Hassan Yusufu for their assistance and support throughout the study. I would like to express my gratitude to Drs. Brandon Kohrt and Lavanya Vasudevan for their feedback and input as my thesis committee members; the co-investigators, Drs. Edith Kwobah, Diana Menya, and Irene Chesire for helping to make this project a reality, and the administrative staff members at the Duke Global Health Institute, namely Lysa Mackeen, Sarah Martin and Michael Russell, for providing support related to fieldwork planning. I would also like to thank my host family in Kenya, the team of community health volunteers and women who participated in this project. Lastly, I would like to thank my husband, family, and friends for supporting me through this and all of my endeavors.

Introduction

Depression is a leading cause of disability in the world (Dewey, 2013). It is the most common mental health condition to affect pregnant women and new mothers worldwide with approximately 10-20 percent of women experiencing it either during pregnancy or in the first 12 months following the delivery of the baby (Shidhaye & Giri, 2014).

Maternal depression is a comprehensive term for a range of depressive disorders that affect expectant mothers and mothers (up to 12 months postpartum). These depressive disorders include prenatal depression, postpartum depression, and postpartum psychosis (Shidhaye & Giri, 2014). Our work focuses on perinatal depression that encompasses prenatal and postpartum depression.

Perinatal depression, along with affecting the mother, the neonate, and the developing infant, also affects the entire family unit. Women with depression are more likely to delay the commencing of prenatal care visits and attend them less frequently than their non-depressed counterparts (Dossett, 2008). New mothers who are depressed have a higher likelihood of engaging in abusive behavior towards their children. The bonding quality between mother and child can, thus, be significantly affected (Dubber, Reck, Müller, & Gawlik, 2014; Vismara, Tambelli, Odorisio, & Marconi, 2014) resulting in negative cognitive and behavioral outcomes in the children and increased risk of

developing emotional disorders (Dossett, 2008).

Perinatal depression is a major public health issue in low and middle-income countries (LAMICs) (Rahman et al., 2012). The prevalence of postpartum depression among women in low-resource settings varies widely but is thought to be higher than 30 percent in rural settings of some developing countries (Villegas, McKay, Dennis, & Ross, 2011). Although the prevalence of depression is similar in pregnant, post-partum and non-pregnant women, the start of new depression is higher during the perinatal period and women of low-income in low to middle-income countries are at an especially high risk (Rochat et al., 2011). While LAMICs comprise 80% of the global population, they only receive less than 20 percent of mental health resources (Patel V & Prince M, 2010).

While there has been increasing awareness, clinical research and public health efforts regarding perinatal depression, it is still a critical health concern for pregnant women and new mothers (Dossett, 2008). A study by Lusskin et al. illustrates that early identification and treatment of perinatal depression will reduce morbidity and mortality for the woman, the child, and the family (Lusskin, Pundiak, & Habib, 2007).

Depression screening has become routine practice in many high-income countries and adapting such tools for use in other contexts is crucial to ensure that they are valid and reliable there (Sweetland, Belkin, & Verdelli, 2014). There are several depression-screening tools that have been validated in pregnant and postpartum

women in the developed world. Among them are the Patient Health Questionnaire-9 (PHQ-9), originally developed to detect depression in primary care populations, and the Edinburgh Postnatal Depression Scale (EPDS), originally designed to detect postpartum depression (Dossett, 2008). There has not, however, been extensive investigation on the cultural validity of instruments to detect perinatal mental disorders in rural settings in developing countries (Hanlon et al., 2008). A systematic review by Tsai et al. highlighted that a major obstacle to improving perinatal mental health in Africa is the lack of locally validated tools to identify probable cases of perinatal depression (Tsai et al., 2013). There is, thus, a critical need to develop and validate novel screening tools for perinatal depression. The aim of our study is two-fold: (i) to validate the Edinburgh Postnatal Depression Scale (EPDS) and the Patient Health Questionnaire-9 (PHQ-9) and (ii) to develop and validate a new screening tool that incorporates questions from existing screening tools with local expressions of depression in rural Kenya.

Methods

Setting and Participants

The study was conducted in Bungoma County, Kenya. Bungoma County is in Western Kenya and was known as Western Province. Over 1.5 million people live in the county, nearly half of whom in poverty (Wiesmann, Kiteme, & Mwangi, 2014).

Participants were recruited for two study activities (i) focus group discussions (FGDs) to facilitate the development of locally relevant screening questions and (ii) administration of the new and two existing screening tools to assess their validity. Eligibility criteria for women to participate in the FGDs were as follows: (i) being at least 18 years of age, (ii) obtaining maternity health services from a specific primary health clinic in the subcounty, Bungoma East. Eligibility criteria for community health volunteers to participate were that they served the specific clinic's catchment area. In Kenya, CHVs are a crucial network of individuals providing services and health education to families who may otherwise not have access to them (Leadership, 2015). They also have a specific focus on neonatal and maternal health ("One Million Community Health Workers," n.d.). Thus, they served as crucial informants with regards to our research topic. Women and CHV focus group discussions were held separately.

To assess the validity of the screening tools, a random sample of pregnant women and new mothers from 27 villages that were either completely or partly situated within a 2-kilometer radius from the clinic was recruited. Eligibility criteria for this portion of the study were as follows: (i) being at least 18 years old, (ii) being in second or third trimester of pregnancy, in the case of pregnant women (iii) being 1 to 6 months

postpartum, in the case of new mothers. We excluded women who had a miscarriage, stillbirth or infant death associated with their most recent pregnancy.

Measures

Screening Survey

We identified 17 commonly used depression screening tools, including the EPDS and PHQ-9 (see Appendix A), and created a database of items from every measure. Items from the “Core” and “Standard” versions of the *Pregnancy Risk Assessment Monitoring System (PRAMS)* and the *Bromley Postnatal Depression Scale (BPDS)* were removed because the measures focused more on patient history instead of depressive symptoms. For instance, the BPDS includes items such as “Did you talk to a psychiatrist because you felt depressed during the first year after the baby was born?” An example PRAMS “Core” item is “Thinking back just before you got pregnant with your new baby, how did you feel about becoming pregnant?” After removing items from these scales, our list of items totaled 365.

We assigned every screening item a “cover term” (e.g., in ill health, temper, unhappy) and reviewed each cover term for exact or approximate duplicates. Out of the initial 365 screening items, we identified 171 unique cover terms. We then created an index card (with English and Kiswahili writing) for each unique cover term in preparation for our focus group discussions. The discussions resulted in the generation

of a 60-item screening tool that comprised of items from the EPDS, the PHQ-9, items from other existing screening tools, and new items generated locally from FGDs. Along with the mental health-specific questions, the survey also consisted of demographic questions from the Phase 6 and 7 Demographic and Health Household and Woman's questionnaires (DHS Program, n.d.).

Edinburgh Postnatal Depression Scale

The most commonly utilized screening instrument for perinatal depression is the 10-item Edinburgh Postnatal Depression Scale (EPDS; Cox, Holden, & Sagovsky, 1987). The EPDS was reported to have a sensitivity, specificity, and positive predictive value of 86, 78, and 73 percent respectively through its first validation study with 84 postnatal women in the United Kingdom. A systematic review of 37 EPDS studies performed between the years 1987 and 2008 illustrated great diversity in sensitivity and specificity values between studies (Gibson, McKenzie-McHarg, Shakespeare, Price, & Gray, 2009).

Tsai et al. (2013) built upon this work with a new systematic review of 25 studies that focused specifically on perinatal depression in Africa. Out of the 25 studies included in the review, 16 used the EPDS. The authors established that the median estimated alpha coefficient for the EPDS was 0.84, the pooled sensitivity and specificity was 0.94 and 0.77 (cut-off ≥ 9). Only 7 out of 12 studies assessed criterion validity used a clinical

interview or semi-structured interview as a gold standard. Also, none of the studies cited in the review were conducted in Kenya.

Patient Health Questionnaire-9

The Patient Health Questionnaire-9 is a depression screening tool that is also often used to evaluate perinatal depression. Two studies have examined Kiswahili translations of the PHQ-9 in Kenya. Omoro et al. (2006) found an association between PHQ-9 scores, TNM stage (Classification of Malignant Tumors), and scores on a quality of life-scale that is cancer-specific. A study by Monahan et al. (2009) asserted that the PHQ-9 was valid in Kenya by comparing scale scores to how patients respond to the question, "In general, how would you rate your overall health right now?" Neither study assessed validity using a criterion reference such as a structured clinical interview.

Criterion Reference: Structured Clinical Interview for DSM-5

Two Kenyan counselors were trained on how to use the Structured Clinical Interview for DSM-5, Research Version (SCID-5-RV; First, Williams, Karg, & Spitzer, 2015) to diagnose depression cases. The SCID-5-RV is customizable, and we chose to administer the non-patient overview, Module A on mood episodes with specifiers, Module B/C for psychotic screening, and Module D for the differential diagnosis of mood disorders.

In order to meet criteria for a current Major Depressive Disorder (MDE) according to the DSM-5 (American Psychiatric Association, 2013), a woman had to experience at least 5 out of 9 symptoms-including depressed mood (A1) or diminished interest or pleasure (A2)- during the same 2-week period within the past 1 month (Criterion A) and report that these symptoms caused clinically significant distress or impairment in functioning (Criterion B). Interviewers looked into all cases in which a general medical condition, substance abuse, or medication could be the etiological factor (Criterion C). Interviewers also used Module B/C to determine if psychotic symptoms were primarily accounted for by a DSM-5 Psychotic Disorder (Criterion D). Interviewers did not examine Criterion E of Module D (i.e., rule out manic or hypomanic episode); therefore, we could only diagnose MDE not Major Depressive Disorder.

Alternate Criterion Reference: Local Diagnosis

Along with using the SCID-5-RV to diagnose depression as defined by the DSM-5, we also asked counselors to use their clinical judgment and women to self-report on the functioning.

Clinical judgment of diagnosis and functioning

Counselors answered the following question to record a 'local' diagnosis that was not related to the DSM criteria: "In your clinical judgment, do you think that this woman is 'depressed'?" Counselors also rated each woman's social and occupational functioning using the Social and Occupational Functioning Assessment Scale (SOFAS) in

the SCID-5-RV. SOFAS scores range from 0 to 100, with 100 representing “superior functioning”.

Client rating of functioning

Counselors asked women to rate their own functioning by selecting on which step of an imaginary 10-step ladder they stood. Women were told that people who were not doing well stood on Step 1, and those who were doing very well stood on Step 10.

Local reference standard: Agreement between counselor and client

If there was agreement between the counselor’s local diagnosis of depression and the woman’s report that she stood on steps 1 through 5 (not doing well), the woman was classified as a ‘local’ case of depression (Bolton, 2001a).

Procedures

Appendix C and D illustrate the cover term card development process and overall study design.

Qualitative study to Develop Potential Screening Items

We developed the set of screening items to be used in the validation study through free listing and card sorting, expert review, and local adaptation. The purpose of this phase of the research study was to determine the content equivalence, content validity, semantic equivalence, and technical equivalence of the screening items (Kohrt et al., 2011).

Free-listing and card sorting activity (content equivalence)

Over the course of a week in June 2015, we conducted free listing and card sorting exercises with 2 groups of pregnant women and new mothers ($n=12$) and 6 groups of CHVs ($n=38$). On average, groups had 6.2 participants ($SD=0.9$). Free listing is an ethnographic research method that results in a list of responses to a single inquiry (Bass et al., 2008, Bolton, 2001, and Kohrt et al., 2011). After a brief discussion about the general maternal health issues in the community, we asked each group to list as many characteristics as they could think of in response to a prompt to describe what depression (“sadness” in Kiswahili) looks like in pregnant women and new mothers in their setting. The student researcher who was fluent in both English and Kiswahili facilitated each discussion and probed for more details throughout the listing procedure. A second researcher took notes and created an index card for each characteristic mentioned by the group.

While the listing exercise was ongoing, a research assistant attempted to match the index cards generated by the group to our set of 171 cover terms (in the patient focus groups we only attempted to match against the 54 most common cover terms). After recording direct matches, all non-matching cards were spread out on the table for the group to review. We asked participants to sort the cards in to four piles according to whether the characteristic described by the non-matching cover term was something

that they observed in the setting: "yes", "no", "maybe", or "no opinion". We also recorded how frequently groups generated local idioms of distress that did not match with the the Western psychiatric cover terms (Kohrt et al., 2011).

Expert review (content validity)

After the completion of the FGDs, a purposive sample of 11 mental health professionals at Moi University Teaching and Referral Hospital (MTRH) was invited by a Kenyan Co-Investigator to review the free listing and card sorting results. Participants included a psychiatrist, two clinical psychologists and mental health ward nurses. The expert review provided detailed information about the symptoms identified in the free lists, including cultural perceptions of depression and mental health. A free listing activity, similar to the one performed in the FGDs, was also conducted with the expert panel. The researchers attempted to match index cards generated by the professionals to the 50 most highly endorsed cover terms from the previous focus groups, and then asked the experts to sort the remaining non-matching cover terms into the four piles.

Item shortlisting and adaptation (semantic and technical equivalence)

For the validity study, we created a 60-item screening tool that we administered to a random sample of pregnant women and new mothers. In addition to the 10 EPDS items and the 9 PHQ-9 items, we selected 28 of the most endorsed cover terms and 5 of the most frequently mentioned local constructs. We developed each of these 33 terms

into a screening item that matched the format and response options of the PHQ-9. For instance, the cover term “temper” became “easily losing your temper” and followed the PHQ-9 prompt: “For the past two weeks, how many times have you been bothered by the following problems?”

Along with including the original 10 EPDS items, we also included 8 revised EPDS items that matched the format and response options of the PHQ-9. First, we rephrased the new set of EPDS items so that women could select how often a particular issue has bothered them in a period of the past two weeks, as opposed to the past one-week. Next, we reworded the three ‘positively worded’ EPDS items to express problems instead. For example, EPDS item 2 is “I have looked forward with enjoyment to things”. This was rephrased to “Not looking forward to things”.

A native Kiswahili speaker translated these 60 items from English to Kiswahili, and a second person with the same linguistic profile blind back-translated to English. A Kiswahili speaker on the research team resolved inconsistencies and conducted cognitive interviews (covered in detail below) using the translated items with a convenience sample of 30 pregnant women and new mothers¹ visiting the clinic. The

¹ 30 pregnant women and new mothers were interviewed. Demographic information was recorded for 14 out of the 30.

team finally made the necessary adaptations to increase comprehension and reviewed the final set of items with a Kenyan psychiatrist (Co-Investigator).

Cognitive interviews

The Kiswahili translation of the 60 items was piloted with a convenience sample of the target population ($n=30$) through cognitive interviews. The purpose of the cognitive interviews was to assess whether or not the questions were understood and if the intended meaning of the questions was maintained. The response options were also piloted during this time. While participants had the option to choose a response, their responses were not recorded and the aim of the activity was to establish comprehension of the questions. All women were asked different questions and the interviews were performed on an individual basis. If 3 women understood a question consecutively, it was placed in a pile noting that the question is well worded and would not require any revisions. However, if a participant did not understand a question, they had the opportunity to provide recommendations for a rephrasing of the question. The reworded question was then asked to subsequent participants. The process was repeated until the phrasing of all questions, original or re-phrased, were understood by 3 consecutive participants.

Quantitative Study to Validate Measures

After the qualitative portion of the study, we recruited a random sample of pregnant women and new mothers to complete three activities: (i) a 60-item screening

consisting of the EPDS, PHQ-9, and our new items; (ii) a re-test within 7 days; and (iii) a semi-structured clinical interview. The objective of this phase of our study was to determine the construct and diagnostic validity of the screening items (Kohrt et al., 2011)

Sampling

According to the 2014 Kenya DHS, around 98 percent of women in Bungoma County who gave birth in the past five years received antenatal care from a skilled provider (Kenya National Bureau of Statistics et al., 2015). Thus, technically, it should have been enough to use the antenatal register as a sampling frame for the population of pregnant women and new mothers residing within the catchment area of the clinic. In practice, though, clinic register can be inaccurate and incomplete. Thus, working with our partner clinic, a list of potentially eligible participants based on a review of the antenatal (ANC), postnatal (PNC), and delivery registers was created prior to conducting two verifying activities.

711 women who were potentially eligible to participate at our study launch in July were identified (i) 308 women in the ANC register who were in their second or third trimester, (ii) 245 women in the ANC register who, based on the expected delivery date that was recorded, should have given birth in the past 1 to 6 months; (iii) 58 women in the PNC register who attended periodic checkups for babies born in the past 1 to 6

months; and (iv) 100 women in the labor and delivery register who delivered a baby at the clinic in the past 1 to 6 months. After de-duplicating the list, 437 names remained.

In the next step, CHVs for each of the 27 villages within a 2-kilometer radius of the clinic reviewed specific lists to confirm completeness and accuracy. CHVs added 134 names and removed 284 names of women who were ineligible, not residing in the village, or duplicated in the list. The survey team then conducted a second verification exercise on the list of 287 names and further removed 11 names, resulting in a sampling frame of 276 women. A random sample of 210 names was drawn from the sampling frame of 276 names.

Screening and retest

Three Bachelor's-level Kenyan enumerators who had studied social work or mental health or had experience administering surveys completed a day of training to learn how to administer the screening surveys using encrypted Android tablets running ODK Collect (version 1.4.5). The enumerators read questions aloud to the participants and recorded the response provided to them. Each day, completed data forms were sent to a secure research server for processing.

The same enumerator returned to the participant to conduct a re-test survey within 7 days.

Semi-structured clinical interview

In order to establish 'caseness', women also participated in a semi-structured clinical interview. Whether a woman took the survey or interview first was randomized based on village. Two Bachelor's-level counselors and two Master's-level counselors were trained on how to complete the SCID-5-RV by a clinical psychologist (Principal Investigator) and a Kenyan psychiatrist (Co-Investigator). First, the team reviewed each section of the interview guide in detail. Next, a research team member, who is also a native speaker of Kiswahili, played the role of a target woman. She followed scripted response sets designed to expose the clinicians to different interview scenarios and the trainees took turns interviewing her. During this process, the non-interviewing trainee also recorded notes and assigned scores. The two clinicians reached 100 percent agreement on the diagnosis across 4 interviews.

The Master's-level counselors supervised the work of the interviewers by reviewing each completed SCID form and related clinical notes. During the first week of interviews, the team met in person every evening to review assessments and clinical notes. Interviewers were trained to report all cases of current suicidal ideation, intent, or attempts, and severe cases of suspected MDE to supervisors for immediate review. If a woman granted permission, the interviewer made a referral to an available counselor in

the nearest town or the psychiatric nurse at the district hospital, when necessary. The research team provided funds for initial care and transportation.

Analysis

Qualitative Study to Develop Potential Screening Items

We created an overall endorsement score for each cover term by taking the average of the values assigned to each term during the health volunteer focus group discussions. If a cover term matched a card generated by a group during the free listing, the term was assigned a value of '4'. Non-matching cover terms were presented for discussion in each group. Non-matching terms were assigned the following values: '3' if the characteristic was endorsed by the group, '2' if it was agreed that the term could be a possible fit in the group, '0' if there was no opinion in the group about that term, and '-1' if the term was rejected by the group. Cover terms were ranked by endorsement score and local items that did not match with cover terms were ranked by the frequency of mention across groups. The expert panel reviewed the results with the research team and made recommendations for the addition and omission of items from the screening tool.

Quantitative Study to Validate Measures

Our strategy was comprised of two stages. First, we examined which items did the best job distinguishing between cases and non-case and then explored which

combination of items optimized scale reliability and classification accuracy. Next, we evaluated the diagnostic validity, construct validity, construct validity, and reliability of the new scale and the two existing scales.

Item Analysis

We utilized an approach to item analysis that was previously used in the development of the General Health Questionnaire (Goldberg, 1972). All items had a 4-point Likert scale ranging from 0 to 3. The original 10 EPDS items had a different set of response options than the rest of the 50 items, but values were assigned to the EPDS response option that ranged from 0 to 3, with a value of 3 representing severe issues. We calculated the proportion of participants who gave the item a value of 2 or 3 for each item. We then subtracted the proportion of endorsement among non-cases from the proportion of endorsement among cases, providing us with a gradient score.

Items that perform better in differentiating between cases and non-cases have higher gradient scores; negative gradient scores illustrate that a large fraction of non-cases endorsed the item compared to cases. Thus, items with a gradient score of 0.05 or less were eradicated from further consideration. Endorsement by non-cases suggests that the item either measures a broader or different construct than depression. Thus, items that were endorsed by more than 25 percent of non-cases were also removed. This process produced a subset of 20 out of the original 60 items for further evaluation.

Item selection

Caseness was evaluated based on the DSM definition. The internal consistency reliability (Cronbach's alpha) and diagnostic validity (area under the curve, sensitivity and specificity, error rate) of all 616,645 possible combinations of these 20 items in scales ranging in size from 2 to 10 items was evaluated. The 'OptimalCutpoints' package (version 1.1-3; Lopez-Raton & Rodriguez-Alvarez, 2015) for R (version 3.2.3; R Core Team, 2015) was used to select the optimal cutpoint for each combination of items that gave equal weight to sensitivity and specificity in the identification of cases. These results were used to select the items that would make up the new depression scale.

Diagnostic validity

The sensitivity, specificity, accuracy, positive and negative likelihood ratios, and area under the curve for the new perinatal depression scale, the revised EPDS, along with that of the original EPDS and PHQ-9 are reported. One incomplete observation was dropped.

Construct validity

Convergent validity was evaluated by calculating the correlation between measures of depression severity and a measure of functioning: participant ladder rating. Discriminant validity was assessed by comparing the mean score of each measure of depression severity by caseness.

Reliability

Internal consistency reliability (Cronbach's alpha) for EPDS, PHQ-9 and the novel scale were estimated for the first administration of the survey. Test-retest reliability for EPDS, PHQ-9 and the new scale was calculated.

Demographics

Cross-tabulations of key demographic variables by maternal status and caseness were ran. In creating the wealth index from the DHS survey questions, DHS guidance and reference values from the Kenya 2008-09 DHS were followed.

Ethics approval

The study protocol was reviewed and approved by the Duke University Institutional Review Board and the Institutional Research and Ethics Committee in Kenya. All members of the research team completed training in research ethics, and all study participants provided informed consent.

Results

Qualitative Study to Develop Potential Screening Items

Participant characteristics

Free listing and card sorting activities were conducted with 2 groups of pregnant women and new mothers ($n=12$) and 6 groups of CHVs ($n=38$). Groups had an average of 6.2 participants ($SD=0.9$). The average age of female participants and CHVs was 28.2 ($SD=3.4$) and 41.4 years ($SD=7.8$), respectively. 55.3 percent of CHVs completed secondary school and 84.2 percent of them were female. Only 33.3 percent of the female client participants had completed secondary school.

Item shortlisting

A total of 153 cards (25.5 cards per group; $SD=3.2$) was generated by the CHV focus group discussions and 58.8 percent of them matched with 1 of the 171 Western cover terms, supporting the notion that several characteristics of depression are universal. The complete set of 171 Western cover terms was discussed with CHVs while only a subset was discussed with the female client groups, thus, we used data exclusively from the CHV group discussions to select a subset of the most endorsed terms to present to the expert panel. The endorsement scores from the CHV groups were summed to create a total endorsement score value for each item across CHV groups. Twenty-four was the highest score a symptom received, meaning that it received a score

of 4 in all 6 CHV groups. We also presented several of the most frequently mentioned local expressions of distress that were not represented in the Western cover term set. The expert panel reviewed and approved the qualitative results. They also added one item: “feeling like you just want to go back to your maternal home”. Ultimately, 60 items were selected to be included in the survey.

Quantitative Study to Validate Measures

Participant Characteristics

Surveys and interviews were conducted with 193 pregnant women and new mothers. Table 1 illustrates participant characteristics by maternal status. About a third of our sample consisted of pregnant women in their second or third trimester (see Appendix F)

Table 1: Participant demographic information

Characteristics	Pregnant women (n=61)	Postpartum (n=132)	All (n=193)
Age , mean (SD)	26.6 (5.5)	27.4 (6.0)	27.1 (5.9)
Parity , mean (SD)	3 (2.5)	3.8 (2.4)	3.5 (2.5)
Education level, frequency (%)			
No schooling	1 (1.6)	0 (0)	1 (0.5)
Primary school	31 (50.8)	78 (59.1)	109 (56.5)
Secondary school	27 (44.3)	44 (33.3)	71 (36.8)
Higher	1 (1.6)	6 (4.6)	7 (3.6)
Vocational training	1 (1.6)	4 (3.0)	5 (2.6)
Works outside home past 7 days , frequency (%)			
Yes	34 (55.7)	90 (68.2)	124 (64.3)
No	27 (44.3)	42 (31.8)	69 (35.8)
Works outside home past 12 months , frequency (%)			
Yes	48 (78.7)	112 (84.9)	160 (82.9)
No	13 (21.3)	20 (15.2)	33 (17.1)
Marital status, frequency (%)			
Currently married	55 (90.2)	114 (86.4)	169 (87.6)
Cohabiting	1 (1.6)	1 (0.8)	2 (1.0)
Not in a union	5 (8.2)	17 (12.9)	22 (11.4)
Wealth index, frequency (%)			
Quintile 1 (Poorest)	11 (18.3)	13 (9.9)	24 (12.5)
Quintile 2	23 (38.3)	72 (54.6)	95 (49.5)
Quintile 3	22 (36.7)	37 (28.0)	59 (30.7)
Quintile 4	4 (6.7)	10 (7.6)	14 (7.3)
Quintile 5 (Richest)	0	0	0

Prevalence of Depression

The diagnostic results by maternal status are illustrated on Table 2. The prevalence of depression ranges from 5.2 to 14.5 depending on the reference standard used. Using the DSM-5 criteria provides the most conservative estimate of 5.2 percent. Alternatively, having the counselors base their diagnosis on clinical judgment and leaving this unspecified increases the prevalence estimate to 14.5 percent. An estimate of 6.2 percent, which is more in line with DSM-5 estimate, results from agreement between local counselors and clients' rating of their own functioning.

Table 2: Diagnostic results by maternal status

Case definition	Non-cases			Pregnant	Cases			All
	Pregnant	Postpartum	Both (%)		Postpartum	Both (%)		
DSM-5 diagnosis	57	125	182 (94.8)	4	6	10 (5.2)	192	
Local, Counselor-Client Agreement	59	121	180 (93.8)	2	10	12 (6.2)	192	
Local, Counselor Only	51	113	164 (85.5)	10	18	28 (14.5)	192	

Item Analysis

The extent to which items discriminated between cases and non-cases is shown on Appendix H. The plot on Appendix H illustrates each item's gradient score for two different case definitions: DSM-5 (triangle) and counselor clinical judgment (circle).

Items to the right of the x-axis were endorsed by a higher percentage of cases compared to non-cases. The higher the gradient score, the bigger the difference in endorsement between cases and non-cases. Along with having high positive gradient scores, items should also have low endorsement by non-cases. Items meeting this criterion have a black fill. We adopted the 20 items with a gradient score >0.05 (based on the local definition) and less than 25 percent endorsement by non-cases (black shapes to the right of the dashed vertical line). The 20 items includes new items such as “Feeling unable to take care of family” and EPDS items such as “Feeling anxious or worried for no good reason”.

Item Selection

A scatterplot of the internal consistency reliability and accuracy of all 616,645 combinations of these 20 screening items in sets of 2 through 10 is shown on Appendix I. The best combination, that maximizes internal consistency reliability and accuracy, is a set of 9 items that include 2 revised EPDS items, 4 items from other Western psychiatric scales, 2 items generated by the focus groups, and 1 item suggested by the expert panel. We refer to this new scale as the Perinatal Depression Screening or PDEPS (see Table 3).

Table 3: Items in Perinatal Depression Screening (PDEPS)

Items in Perinatal Depression Screening (PDEPS)

From existing screening tools

1. Feeling hopeless about the future
2. Feeling that you are low when compared to other people
3. Thinking that there are problems with your mind
4. Thinking it would be better if you had never been born

Revised EPDS items

5. Feeling anxious or worried for no good reason
6. Crying because of sadness

From free listing activity (local)

7. Feeling unable to take care of your children or family
8. Having problems with partner or other loved one

From expert panel review

9. Feeling like you just want to go back to your maternal home

Measure validation

Diagnostic validity

Using a cutoff of 13, the PDEPS correctly classifies 90 percent of participants according to the DSM-5 criteria, with a sensitivity and specificity of 0.90 and 0.90, respectively. A PDEPS score at or above the cutoff is 9 times for likely among women diagnosed with depression than women without a diagnosis (DSM-5) (see Table 4).

The PDEPS does better than the PHQ-9 and both versions of the EPDS in terms of classification accuracy (DSM-5). The optimal cutoff scores are higher than what most previous work would recommend at 16 and 15 for the original EPDS and PHQ-9,

respectively. The estimates of sensitivity and specificity for the EPDS and PHQ-9, while lower compared to the PDEPS, are within range of what is reported in other studies.

Table 4: Diagnostic validity

Scale	Cut	Sensitivity	Specificity	Accuracy	LRP	LRN	AUC
EPDS, original	≥16	0.70	0.72	0.72	2.50	0.42	0.80
EPDS, revised	≥13	0.60	0.73	0.72	2.23	0.55	0.80
PHQ-9	≥15	0.70	0.74	0.73	2.65	0.41	0.79
PDEPS	≥13	0.90	0.90	0.90	8.62	0.11	0.89

Note. Accuracy: 1-error rate. LRP= likelihood ratio (positive). LRN= likelihood ratio (negative). AUC=area under the receiver operating characteristic curve.

Construct validity

The functional impairment scale scores categories are derived from responses of a patient functioning ladder where 1 represents people who are struggling and 10 refers to those who are doing very well. It is known that depression has a negative effect on functioning (Bass, Ryder, Lammers, Mukaba, & Bolton, 2008), thus, convergent validity was evaluated by an association of high dysfunction and increased depression severity. We hypothesized that women with low functioning scores (indicating high function impairment) have higher mean depression scores across all depression screening measures. Conversely, those with high functioning scores (indicating low function impairment) also have lower mean depression scores across all depression screening assessments. There was consistently higher depression severity scores associated with

higher functional impairment across all scales indicating convergent validity (see Table 5).

Table 5: Mean scale scores for cases and non-cases to evaluate convergent validity of PDEPS and existing scales

	Functional impairment scale scores, categorized			
	1-3 (n=6)	4-5 (n=58)	6-7 (n=93)	8-10 (n=33)
EPDS Original, mean (SD)	11.5 (6.3)	14.1 (6.4)	12.2 (5.3)	10 (5.7)
EPDS Revised, mean (SD)	10.7 (5.3)	7.7 (5.5)	7.1 (4.9)	3.8 (4.7)
PHQ-9, mean (SD)	13.5 (5.5)	12.1(6.2)	10.2 (6.2)	7.4 (6.4)
PDEPS, mean (SD)	8.3 (4.2)	8.2 (5.9)	6.7 (5.0)	4.4 (5.0)

To test discriminant validity, the mean syndrome severity scores between ‘cases’ and ‘non-cases’ across scales was compared. Comparing the total depression symptom scores for PDEPS amongst the two groups, the mean value for the cases is more than double that of the non-cases, which is statistically significant (see Table 6). The cases have higher mean scores than the non-cases across all scale combinations.

Table 6: Mean scale scores for cases and non-cases to evaluate discriminant validity of PDEPS and existing scales

Scale	Cases (n=10)	Non-cases (n=181)	Diff (%)
EPDS Original, mean (SD)	18.9 (6.2)	12.0 (5.6)	56.9**
EPDS Revised, mean (SD)	15.1 (4.3)	8.5 (6.4)	77.6***
PHQ-9, mean (SD)	16.3 (4.1)	10.0 (6.3)	62.6***
PDEPS, mean (SD)	13.6 (3.0)	6.3 (5.2)	115.8***

Note. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.

Reliability

Table 7 presents estimates of internal consistency reliability and test-retest reliability. Each scale illustrates acceptable internal consistency reliability with Cronbach's alpha scores ranging from 0.77 to 0.83 (Nunnally, 1978).

The correlations between the scale scores from the first survey and the retest range from 0.59 to 0.63 suggesting sufficient reliability.

Table 7: Reliability of PDEPS and existing scales

Scale	Internal consistency		Test-retest	
	<i>N</i>	Alpha	<i>n</i>	<i>r</i>
EPDS, Original	193	0.78	95	0.57
EPDS, Revised	193	0.83	95	0.64
PDEPS	193	0.77	95	0.67
PHQ-9	193	0.81	95	0.62

Discussion

This study demonstrated that the EPDS and PHQ-9 screening tools have acceptable sensitivity and specificity to detect depression among pregnant women and new mothers in Kenya. The diagnostic validity for the EPDS are at the low end of what is cited in other studies done in Africa and our recommended cutoff of ≥ 16 is higher than what other studies utilize. Our results confirm that the EPDS and PHQ-9 are valid tools for this setting. They may not be optimal for use there though.

The Perinatal Depression Screening, or PDEPS was created through a process of free listing, card sorting, and item analysis. It does better on all metrics of classification accuracy when considering DSM-5 caseness. We cannot draw strong conclusions about the diagnostic validity of the PDEPS to detect depression during the antenatal period because too few pregnant women were recruited. Overall classification accuracy was high, however. Our results show that the PDEPS outperforms the EPDS and PHQ-9 in terms of diagnostic accuracy during the postpartum period. We are able to state this with confidence, as our postpartum sample was twice as large as our prenatal one.

The PDEPS was developed through an approach that integrated both Western psychiatric concepts with locally anchored idioms of distress. We drew inspiration from several previous studies that took a similar approach of combining Western and local concepts (e.g. Bass et al., 2008; Kaaya, Lee, Mbwambo, Smith-Fawzi, & Leshabari, 2008; Nhiwatiwa, Patel, & Acuda, 1998; Patel, Simunyu, Gwanzura, Lewis, & Mann, 1997).

Our results can most specifically be compared to the work by Bass et al. (2008) who used similar methodology in the Democratic Republic of Congo to develop and validate a locally-derived measure of postpartum depression (*Maladi ya Souci*, a syndrome of worry) that had some EPDS items. The researchers found that their 16-item local syndrome scale was similar to the shorter version of the EPDS and another Western scale in terms of diagnostic accuracy.

Unlike Bass et al. (2008) who found that their locally-grounded scale included all of the diagnostic symptom categories for MDD, we found that the best combination of questions on the PDEPS somewhat deviates from DSM-5 criteria. For example, an item that distinguished between a case and non-case well, “Felling like you just want to go back to your maternal home”, was suggested by the expert panel of Kenyan mental health practitioners who are trained in the Western model but whose practice is informed by local customs. This particular custom refers to the tradition that women often leave their maternal home upon marriage to live in their husband’s village. Given the culturally rooted nature of this question, some adaptation might be required if the PDEPS is to be used in settings where such is not a custom. Other PDEPS questions like “feeling hopeless”, “crying because of sadness”, and “feeling anxious or worried for no good reason” highlight the universality of depression symptoms.

The study by Bass et al. (2008) is also an interesting comparison because there was no use of standard clinical interviews to determine caseness. Instead, they

considered a woman to be a 'case' if a key informant identified her as suffering from the local syndrome and if the woman self-identified as having the syndrome. In our study, along with using the SCID-5-RV as a gold standard, we also examined 'local' cases defined by agreement between a counselor's clinical judgment and the woman's self-report of poor functioning. To our knowledge, ours is the first study to assess Western and local case defining approaches.

Both definitions of caseness interestingly lead to similar overall estimates of prevalence: 5.2 percent using DSM-5 criteria and 6.2 percent using local definition based on agreement between counselor clinical judgment and client self-rating. These results deviate from other published prevalence estimates. For example, a systematic review by Villegas et al. (2011) reported an estimate of 31.3 percent in developing countries (95%CI 21.3 to 43.5). Another systematic review of common perinatal mental disorders in low- and lower-middle-income countries estimated a prevalence of 18.6 percent (95%CI 18.0 to 19.2; Fisher et al., 2012). A third systematic review limited to studies in Africa reported a prevalence of depression of 11.3 percent during the antenatal period (95%CI 9.5 to 13.1) and 18.3 percent during the postnatal period (95%CI 17.5 to 19.1; Sawyer, Ayers, & Smith, 2010).

When we consider only counselor clinical judgment, our prevalence estimate increases to 14.5 percent. This might be most comparable to the Fisher et al. (2012) estimate of 18.6 percent for common perinatal mental disorders, assuming that

counselors in our study might have considered a broader range of symptoms than strictly depression when labeling cases relying on their clinical judgment.

Our study employed ethnographic methods that stress providing participants with very little information about interviewer expectations. This is essential in order to ensure not to influence the responses provided by the participants by the interviewer and how questions are worded (Bolton, 2001b). With this point considered, it is interesting how several of the symptoms mentioned during the free listing activities matched with depression symptom descriptions in the DSM-5 highlighting the universal nature of depression symptoms across cultures.

It is also noteworthy that out of the 5 locally derived questions (see Appendix G), 4 of them (except feeling stressed) are socially focused and involve not just one's perceptions of self but how one interacts with others. Existing Western-centric depression screening tools may not have such socially inclined questions, as social manifestations of depression may not be as prominent in Western culture. While 'feeling like you want to be alone' is a manifestation of depression in rural Kenya, wanting to be alone in Western setting may simply be a sign of withdrawal or preoccupation (Betancourt, Spielman, Onyango, & Bolton, 2009).

Study strengths and limitations

The strengths of the study included the fact that we culturally-adapted the depression-screening tool prior to administering it using strong ethnographic research

methods. Along with holding focus group discussions with pregnant women, new mothers and CHVs, we also held an expert panel review with mental health practitioners. While the former group was able to provide a more first hand experience of the issue, the latter was able to confirm the findings based on the wide variety of cases that they encounter.

An additional strength of our study is that we had a systematic approach to establishing a sample frame which not only involved consolidating data from ANC, PNC, and delivery logs but also having multiple rounds of verification by CHVs- the primary source of health information for a majority of women in the area thus key individuals to be involved in this process. A majority of studies obtain their sampling frame from individuals visiting a clinic alone. Through the approach to develop a sampling frame in our study, we were able to obtain a representative community sample.

While our study had points that strengthened it, it should be considered with the following limitations in mind. Our study had two limitations. First, our sample size limited us in being able to precisely disaggregate results by pregnancy status.

Second, our research project was carried out in a catchment area of a 2-kilometer radius around a particular primary health care facility. While our results are applicable to the women in that area, they may not be generalizable to the wider rural community in Kenya.

Conclusion

The EPDS and PHQ-9 are valid and reliable screening tools for perinatal depression in rural Western Kenya, but a new 9-item locally developed tool called the Perinatal Depression Scale (PDEPS) that incorporates both Western psychiatric concepts and local idioms of distress may be a better alternative. The prevalence of depression in this region, at less than 10 percent, appears to be lower than other published estimates for countries in Africa and other low-income countries. Our findings point to the importance of adapting and validating depression-screening tools. Additional research is needed to confirm these findings.

Appendix A

Measures of depression

Measures	Items
SCL-90	90
General Health Questionnaire	60
Pregnancy Risk Assessment Monitoring System Core	38
Response Inventory for Stressful Life Events	36
Hopkins Symptoms Checklist for Depression	24
Pitt Depression Scale	24
Becks Depression and Anxiety Scales	21
Center for Epidemiological Studies Depression Scale	20
Zung's Self Rating Depression Scale	20
Self Reporting Questionnaire	20
Bromley Postnatal Depression Scale	17
Pregnancy Risk Assessment Monitoring System	17
Hospital Anxiety and Depression Scale	14
Edinburgh Postnatal Depression Scale	10
Kessler Mental Health Distress Scale	10
Patient Health Questionnaire-9	9
Pregnancy Risk Assessment Monitoring System 6	6

Note: We removed items included in the "Core" and "Standard" versions of the Pregnancy Risk Assessment Monitoring System (PRAMS) and the Bromley Postnatal Depression Scale (BPDS) because the scales assess a patient's history rather than her depressive symptoms. For instance, the BPDS includes items like "Did you talk to a psychiatrist because you felt depressed during the first year after the baby was born?" An example PRAMS "Core" question is "Thinking back just before you got pregnant with your new baby, how did you feel about becoming pregnant?"

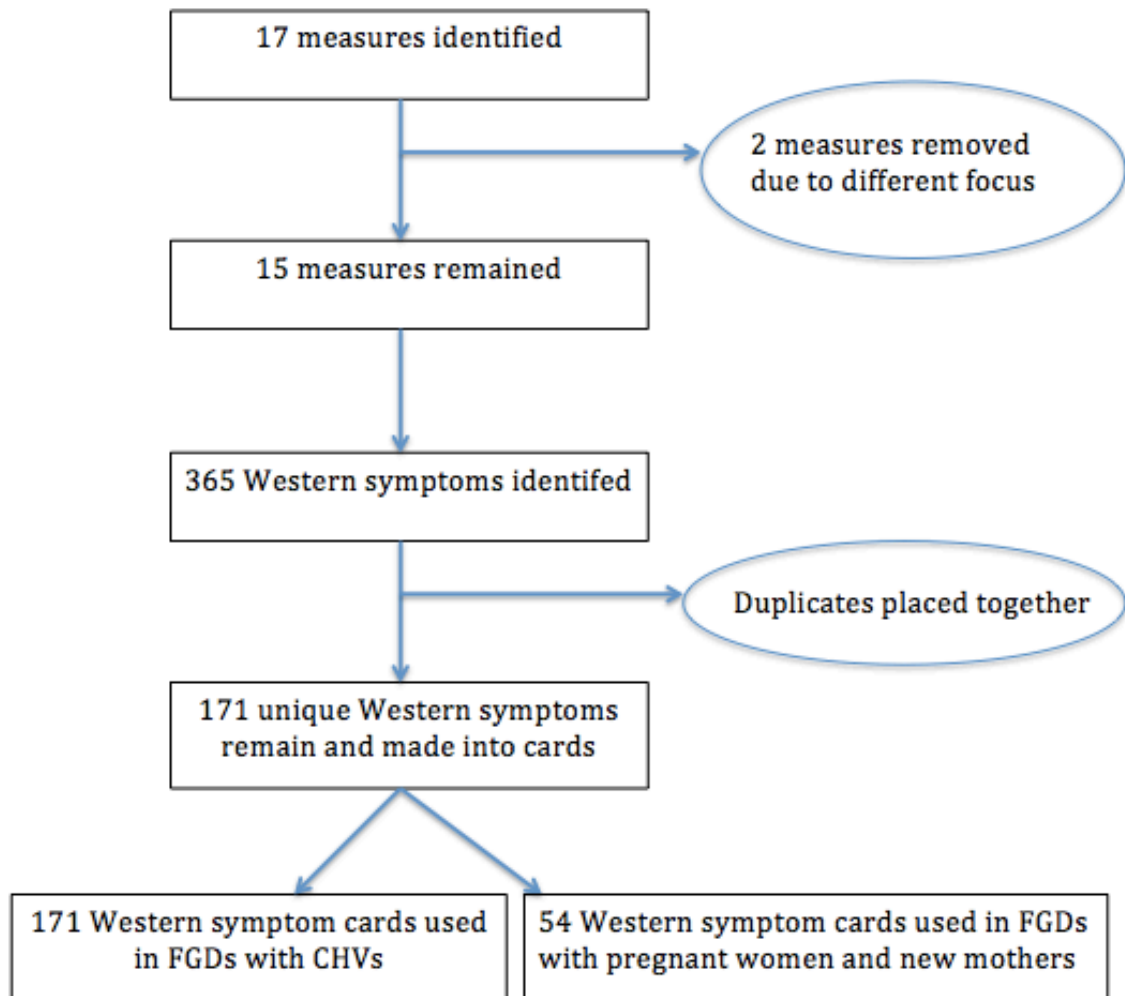
Appendix B

Top 25 most endorsed cover terms

Concept	Score
In ill health	24
Temper	24
Unhappy	23
Change in appetite	22
Depressed	22
Poor appetite	22
Unwanted thoughts, words, ideas in mind	22
Losing weight	21
Not care about appearance	21
Talk less	21
Think something wrong with mind	21
Irritable	20
Tired/low energy	20
Uncomfortable feelings in stomach	20
Loss interest/pleasure in sex	19
Loss of sleep	19
Bothered by shouting or throwing things	18
Couldn't overcome difficulties	18
Difficulty concentrating	18
Feel uneasy in crowds	18
Feel unlike normal self	18
Feeling weak in parts of body	18
Headaches	18
Hot or cold spells	18
Mind not clear	18

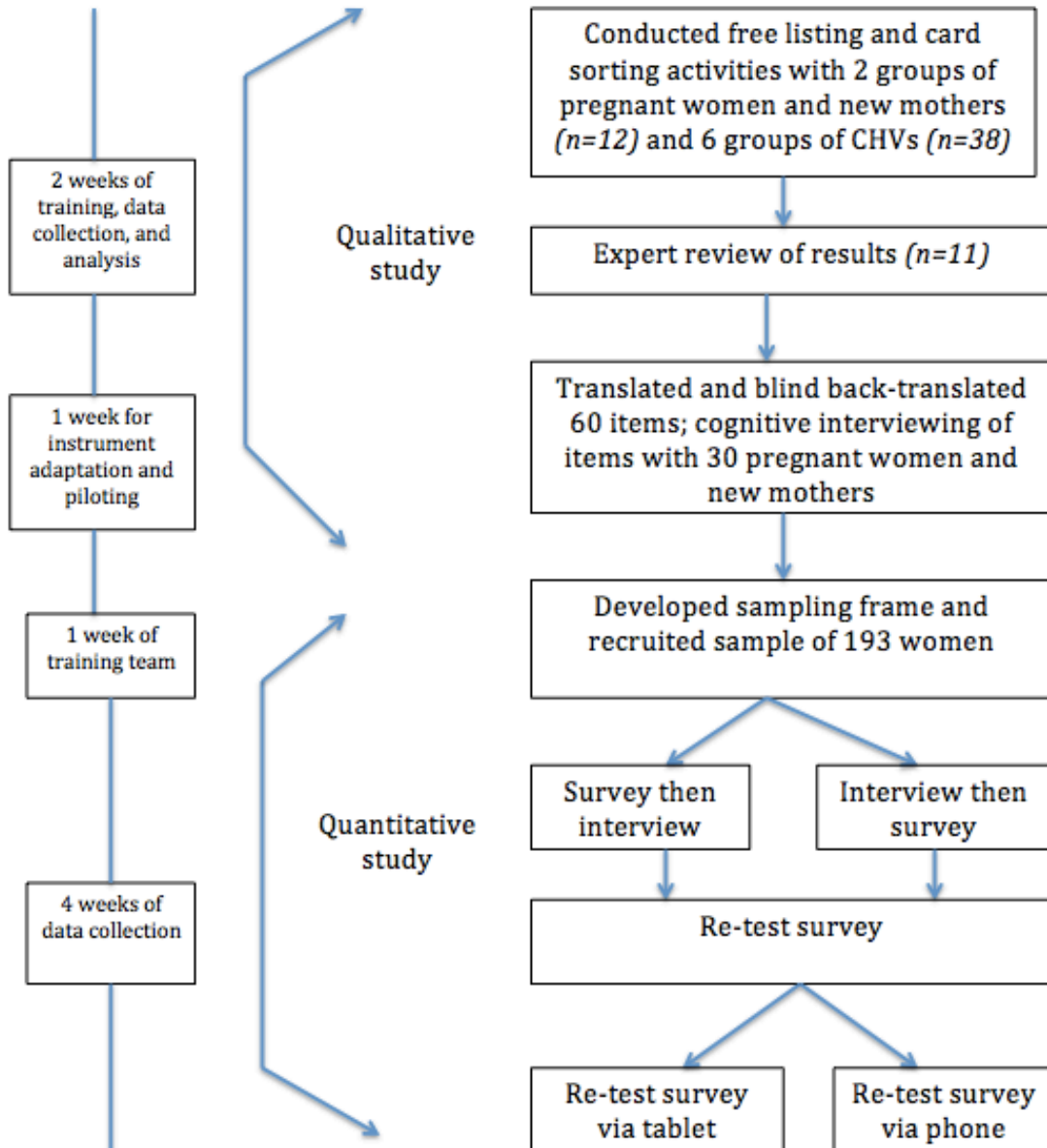
Appendix C

Cover term card development process



Appendix D

Eight-week study schema



Appendix E

Original and Revised EPDS questions

Original EPDS question	Revised EPDS question
1. I have been able to laugh and see the funny side of things	1. <i>Feeling unable to laugh or have fun</i>
2. I have looked forward with enjoyment to things	2. <i>Not looking forward to things</i>
3. I have blamed myself unnecessarily when things went wrong	3. <i>Blaming yourself when things go wrong</i>
4. I have been anxious or worried for no good reason	4. <i>Feeling anxious or worried for no good reason</i>
5. I have felt scared or panicky for no very good reason	
6. Things have been getting on top of me	6. <i>Feeling overwhelmed</i>
7. I have been so unhappy that I have had difficulty sleeping	
8. I have felt sad or miserable	8. <i>Feeling sad or miserable</i>
9. I have been so unhappy that I have been crying	9. <i>Crying because of sadness</i>
10. Thoughts of harming myself have occurred to me	10. <i>Thoughts of harming yourself</i>

Appendix F

Number of ANC visits based on number of months pregnant

Number of months pregnant	Number of ANC visits, frequency (%)					
	0	1	2	3	4	Total
2-4	6 (54.55)	5 (45.45)	0	0	0	11 (100.00)
5-7	4 (14.81)	10 (37.04)	5 (18.52)	6 (22.22)	2 (7.41)	27 (100.00)
8	0	0	4 (19.05)	11 (52.38)	6 (28.57)	21 (100.00)
>8	0	0	0	1 (50.00)	1 (50.00)	2 (100.00)

Note: A woman who is 2-4 months pregnant is expected to have had at least 1 ANC visit. A woman who is 5-7 months pregnant is expected to have had at least 2 ANC visits. A woman who is 8 months pregnant is expected to have had at least 3 ANC visits. A woman who is more than 8 months pregnant is expected to have had at least 4 ANC visits.

Appendix G

Five locally derived questions

Five locally-derived questions

Feeling unable to take care of your children or family

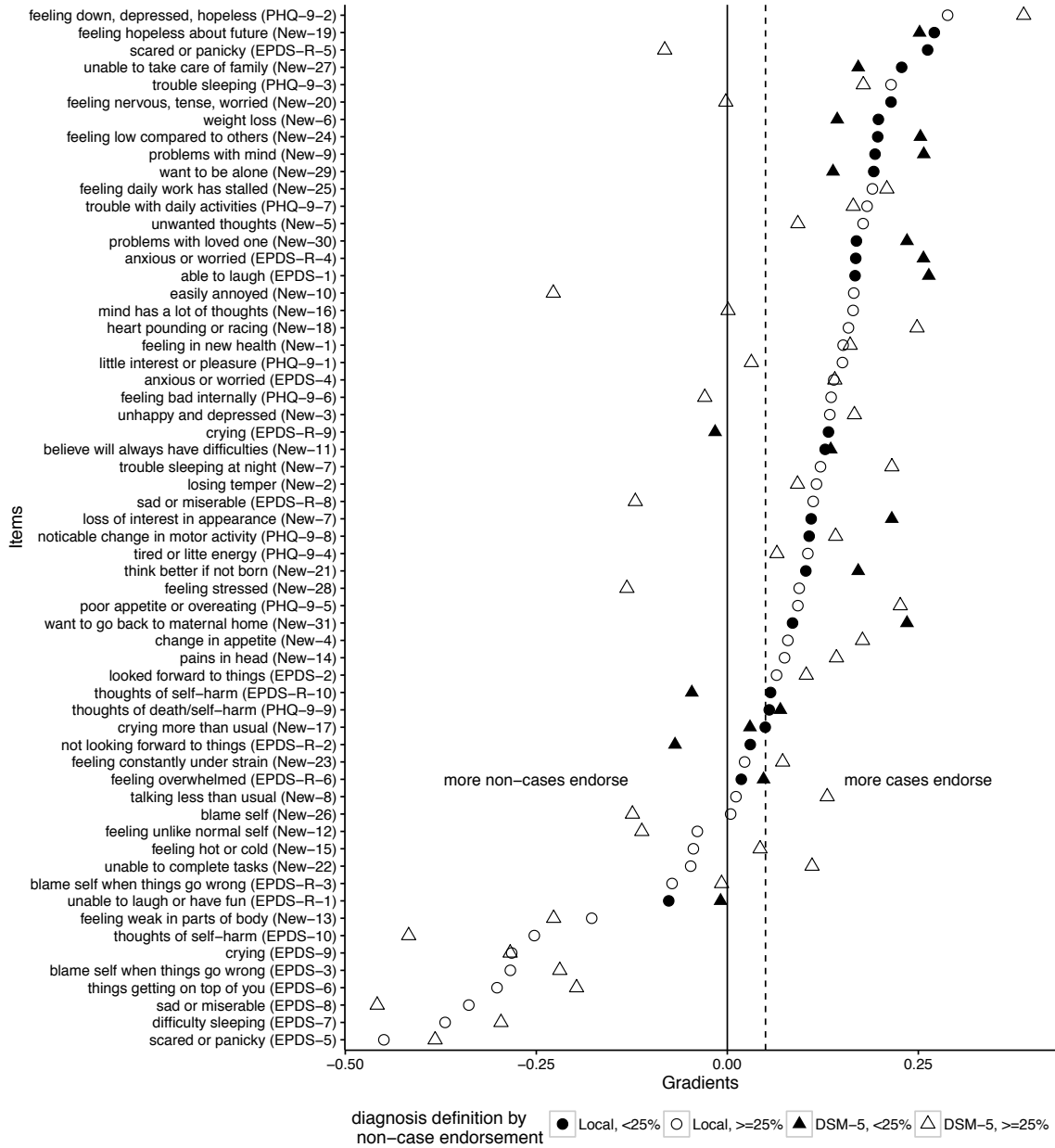
Feeling stressed

Feeling like you want to be alone

Having problems with partner or other loved one

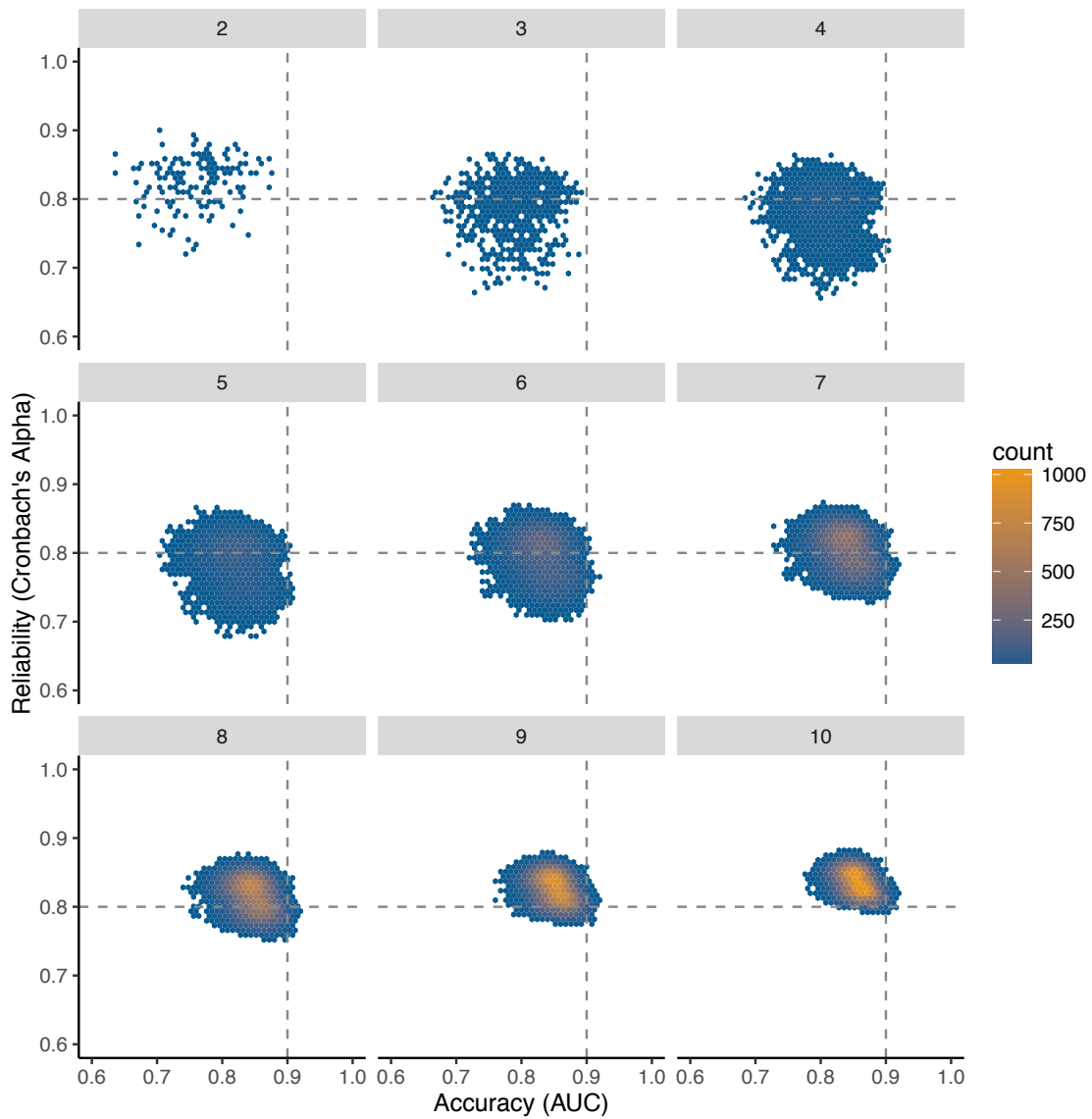
Feeling like you just want to go back to your maternal home

Appendix H



Note. Item gradients illustrating discrimination between cases and non-cases, sorted by case definition

Appendix I



Note. High density scatterplot of the internal consistency reliability and accuracy of all 616,645 combinations of 20 screening items in sets of 2 through 10. DSM-5 definition of depression.

References

- American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders: DSM-5* (5th ed.). Washington, DC: APA.
- Bass, J. K., Ryder, R. W., Lammers, M.-C., Mukaba, T. N., & Bolton, P. A. (2008). Postpartum depression in Kinshasa, Democratic Republic of Congo: Validation of a concept using a mixed-methods cross-cultural approach. *Tropical Medicine & International Health*, *13*(12), 1534–1542. <http://doi.org/10.1111/j.1365-3156.2008.02160.x>
- Betancourt, T. S., Spielman, L., Onyango, G., & Bolton, P. (2009). Psychosocial Problems of War-Affected Youth in Northern Uganda: A Qualitative Study. *Transcultural Psychiatry*, *46*(2), 238–256. <http://doi.org/10.1177/1363461509105815>
- Bolton, P. (2001a). Cross-cultural validity and reliability testing of a standard psychiatric assessment instrument without a gold standard. *The Journal of Nervous and Mental Disease*, *189*(4), 238–242.
- Bolton, P. (2001b). Local perceptions of the mental health effects of the Rwandan genocide. *The Journal of Nervous and Mental Disease*, *189*(4), 243–248.
- Cox, J. L., Holden, J. M., & Sagovsky, R. (1987). Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. *The British Journal of Psychiatry*, *150*(6), 782–786.
- Dewey, C. (2013, November 7). A stunning map of depression rates around the world. *The Washington Post*. Retrieved from <https://www.washingtonpost.com/news/worldviews/wp/2013/11/07/a-stunning-map-of-depression-rates-around-the-world/>
- DHS Program. (n.d.). DHS Model Questionnaires. Retrieved March 7, 2016, from <http://dhsprogram.com/What-We-Do/Survey-Types/DHS-Questionnaires.cfm>
- Dossett, E. C. (2008). Perinatal Depression. *Obstetrics and Gynecology Clinics of North America*, *35*(3), 419–434. <http://doi.org/10.1016/j.ogc.2008.04.004>

- Dubber, S., Reck, C., Müller, M., & Gawlik, S. (2014). Postpartum bonding: the role of perinatal depression, anxiety and maternal–fetal bonding during pregnancy. *Archives of Women's Mental Health, 18*(2), 187–195. <http://doi.org/10.1007/s00737-014-0445-4>
- First, M. B., Williams, J. B. W., Karg, R. S., & Spitzer, R. (2015). *Structured Clinical Interview for DSM-5 – Research Version*. Arlington, VA: American Psychiatric Association.
- Fisher, J., Mello, M. C. de, Patel, V., Rahman, A., Tran, T., Holton, S., & Holmes, W. (2012). Prevalence and determinants of common perinatal mental disorders in women in low-and lower-middle-income countries: a systematic review. *Bulletin of the World Health Organization, 90*(2), 139–149.
- Gibson, J., McKenzie-McHarg, K., Shakespeare, J., Price, J., & Gray, R. (2009). A systematic review of studies validating the Edinburgh Postnatal Depression Scale in antepartum and postpartum women. *Acta Psychiatrica Scandinavica, 119*(5), 350–364. <http://doi.org/10.1111/j.1600-0447.2009.01363.x>
- Goldberg, D. P. (1972). *The detection of psychiatric illness by questionnaire: A technique for the identification and assessment of non-psychotic psychiatric illness*. Oxford, England: Oxford U. Press.
- Hanlon, C., Medhin, G., Alem, A., Araya, M., Abdulahi, A., Hughes, M., ... Prince, M. (2008). Detecting perinatal common mental disorders in Ethiopia: validation of the self-reporting questionnaire and Edinburgh Postnatal Depression Scale. *Journal of Affective Disorders, 108*(3), 251–262. <http://doi.org/10.1016/j.jad.2007.10.023>
- Kaaya, S. F., Lee, B., Mbwambo, J. K., Smith-Fawzi, M. C., & Leshabari, M. T. (2008). Detecting depressive disorder with a 19-item local instrument in Tanzania. *The International Journal of Social Psychiatry, 54*(1), 21–33.
- Kenya National Bureau of Statistics, Kenya Ministry of Health, National AIDS Control Council, Kenya Medical Research Institute, National Council for Population and Development, & ICF International. (2015). *Kenya Demographic and Health Survey 2014*. Nairobi, Kenya: KNBS.
- Kohrt, B. A., Jordans, M. J., Tol, W. A., Luitel, N. P., Maharjan, S. M., & Upadhyaya, N. (2011). Validation of cross-cultural child mental health and psychosocial research

instruments: adapting the Depression Self-Rating Scale and Child PTSD Symptom Scale in Nepal. *BMC Psychiatry*, 11(1), 127.

Leadership, V. (2015, November 9). Community Health Volunteers in Kenya. Retrieved from <http://vision2017.csis.org/kenya-community-health-volunteers/>

Lopez-Raton, M., & Rodriguez-Alvarez, M. X. (2015). OptimalCutpoints (Version 1.1-3) [R].

Lusskin, S. I., Pundiak, T. M., & Habib, S. M. (2007). Perinatal Depression: Hiding in Plain Sight. *Canadian Journal of Psychiatry*, 52(8), 479–88.

Monahan, P. O., Shacham, E., Reece, M., Kroenke, K., Ong'or, W. O., Omollo, O., ... Ojwang, C. (2009). Validity/reliability of PHQ-9 and PHQ-2 depression scales among adults living with HIV/AIDS in western Kenya. *Journal of General Internal Medicine*, 24(2), 189–197.

Nhiwatiwa, S., Patel, V., & Acuda, W. (1998). Predicting postnatal mental disorder with a screening questionnaire: a prospective cohort study from Zimbabwe. *Journal of Epidemiology and Community Health*, 52(4), 262–266.

Nunnally, J. C. (1978). *Psychometric theory*. McGraw-Hill.

Omoro, S. A. O., Fann, J. R., Weymuller, E. A., Macharia, I. M., & Yueh, B. (2006). Swahili translation and validation of the Patient Health Questionnaire-9 depression scale in the Kenyan head and neck cancer patient population. *The International Journal of Psychiatry in Medicine*, 36(3), 367–381.

One Million Community Health Workers | Update from the Field: Kenya Hosts Stakeholders Meeting on CHW Planning. (n.d.). Retrieved from <http://1millionhealthworkers.org/2014/06/10/update-from-the-field-kenya-hosts-stakeholders-meeting-on-chw-planning/>

Patel V, & Prince M. (2010). Global mental health: A new global health field comes of age. *JAMA*, 303(19), 1976–1977. <http://doi.org/10.1001/jama.2010.616>

Patel, V., Simunyu, E., Gwanzura, F., Lewis, G., & Mann, A. (1997). The Shona Symptom Questionnaire: the development of an indigenous measure of common mental disorders in Harare. *Acta Psychiatrica Scandinavica*, 95(6), 469–475.

- Rahman, A., Sikander, S., Malik, A., Ahmed, I., Tomenson, B., & Creed, F. (2012). Effective treatment of perinatal depression for women in debt and lacking financial empowerment in a low-income country. *The British Journal of Psychiatry*, 201(6), 451–457. <http://doi.org/10.1192/bjp.bp.112.109207>
- {R Core Team}. (2015). *R: A Language and Environment for Statistical Computing*. {Vienna, Austria}: {R Foundation for Statistical Computing}. Retrieved from <https://www.R-project.org/>
- Rochat, T. J., Tomlinson, M., Bärnighausen, T., Newell, M.-L., & Stein, A. (2011). The prevalence and clinical presentation of antenatal depression in rural South Africa. *Journal of Affective Disorders*, 135(1-3), 362–373. <http://doi.org/10.1016/j.jad.2011.08.011>
- Sawyer, A., Ayers, S., & Smith, H. (2010). Pre- and postnatal psychological wellbeing in Africa: A systematic review. *Journal of Affective Disorders*, 123(1–3), 17–29. <http://doi.org/10.1016/j.jad.2009.06.027>
- Shidhaye, P., & Giri, P. (2014). Maternal depression: a hidden burden in developing countries. *Annals of Medical and Health Sciences Research*, 4(4), 463.
- Sweetland, A. C., Belkin, G. S., & Verdeli, H. (2014). Measuring depression and anxiety in Sub-Saharan Africa. *Depression and Anxiety*, 31(3), 223–232. <http://doi.org/10.1002/da.22142>
- Tsai, A. C., Scott, J. A., Hung, K. J., Zhu, J. Q., Matthews, L. T., Psaros, C., & Tomlinson, M. (2013). Reliability and Validity of Instruments for Assessing Perinatal Depression in African Settings: Systematic Review and Meta-Analysis: e82521. *PLoS One*, 8(12), e82521. <http://doi.org/http://dx.doi.org.proxy.lib.duke.edu/10.1371/journal.pone.0082521>
- Villegas, L., McKay, K., Dennis, C.-L., & Ross, L. E. (2011). Postpartum depression among rural women from developed and developing countries: A systematic review. *The Journal of Rural Health*, 27(3), 278–288.
- Vismara, L., Tambelli, R., Odorisio, F., & Marconi, P. (2014). EPA-1274 – Maternal representations and mother-child interaction in the perinatal period: effects of depression and anxiety comorbidity. *European Psychiatry*, 29, Supplement 1, 1. [http://doi.org/10.1016/S0924-9338\(14\)78505-0](http://doi.org/10.1016/S0924-9338(14)78505-0)

Wiesmann, U., Kiteme, B., & Mwangi, Z. (2014). *Socio-Economic Atlas of Kenya: Depicting the National Population Census by County and Sub-Location*. Nairobi: KNBS.