



Published in final edited form as:

J Affect Disord. 2023 March 01; 324: 637–644. doi:10.1016/j.jad.2022.12.101.

Comparing depression screening tools (CESD-10, EPDS, PHQ-9, and PHQ-2) for diagnostic performance and epidemiologic associations among postpartum Kenyan women: Implications for research and practice

Anna Larsen, MPH¹, Jillian Pintye, PhD², Ben Odhiambo, BS³, Nancy Mwangeli, BS³, Mary M. Marwa, BS³, Salphine Watoyi, BS³, John Kinuthia, MBBS³, Felix Abuna, BS³, Laurén Gomez, MPH², Julia Dettinger, MPH², Amritha Bhat, MD⁴, Grace John-Stewart, MD^{1,2,5,6}

¹University of Washington, Department of Epidemiology

²University of Washington Department of Global Health

³University of Nairobi/Kenyatta National Hospital

⁴University of Washington, Department of Psychiatry and Behavioral Sciences

⁵University of Washington, Department of Medicine

⁶University of Washington, Department of Pediatrics

Abstract

Corresponding author: Anna Larsen, MPH; University of Washington, Department of Epidemiology, 3980 15th Ave NE, Box 351619, Seattle, WA 98195; Tel: +1-206-334-2783, annalar@uw.edu.

Author's statement:

Contribution to authorship:

AL-conceptualized the idea for the article, performed analyses, wrote the first draft, conducted the edits for the drafts, drafted the figures and tables, reviewed and approved final draft.

JP-reviewed the data in the figures and tables, reviewed drafts and provided substantial edits and contributed towards final draft, reviewed and approved final draft.

BO- managed data collection and study operations, reviewed drafts and provided substantial edits and contributed towards final draft, reviewed and approved final draft.

NM- managed study operations, reviewed drafts and provided substantial edits and contributed towards final draft, reviewed and approved final draft.

MM- performed data management, reviewed drafts and provided substantial edits and contributed towards final draft, reviewed and approved final draft.

SW- performed data management, reviewed drafts and provided substantial edits and contributed towards final draft, reviewed and approved final draft.

JK- directed study operations, reviewed drafts and provided substantial edits and contributed towards final draft, reviewed and approved final draft.

FA- managed study operations and study team, reviewed drafts and provided substantial edits and contributed towards final draft, reviewed and approved final draft.

LG- reviewed drafts and provided substantial edits and contributed towards final draft, reviewed and approved final draft.

JD- reviewed drafts and provided substantial edits and contributed towards final draft, reviewed and approved final draft.

GJS- conceptualized the idea for the article, reviewed the data in the figures and tables, reviewed all drafts and provided substantial edits and contributed towards final draft, reviewed and approved final draft.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Background: Identifying optimal depression screening tools for use in maternal health clinics could improve maternal and infant health. We compared four tools for diagnostic performance and epidemiologic associations.

Methods: This study was nested in a cluster-randomized trial in Kenya. Women in 20 maternal health clinics were evaluated at 6 weeks postpartum with Center for Epidemiologic Studies Depression Scale (CESD-10), Edinburgh Postnatal Depression Scale (EPDS), Patient Health Questionnaire-9 and -2 (PHQ-9, PHQ-2) for moderate-to-severe depressive symptoms (MSD) [CESD-10 10, EPDS 13, PHQ-9 10, or PHQ-2 3]. We assessed area under the curve (AUC) per scale (CESD-10, EPDS) against probable major depressive disorder (MDD) using the PHQ-9 scoring algorithm. Associations between MSD and intimate partner violence (IPV) were compared between scales.

Results: Among 3605 women, median age was 24 and 10% experienced IPV. Prevalence of MSD symptoms varied by tool: 13% CESD-10, 9% EPDS, 5% PHQ-2, 3% PHQ-9. Compared to probable MDD, the CESD-10 (AUC:0.82) had higher AUC than the EPDS (AUC:0.75). IPV was associated with MSD using all scales: EPDS (RR:2.5, 95%CI:1.7–3.7), PHQ-2 (RR:2.3, 95%CI:1.6–3.4), CESD-10 (RR:1.9, 95%CI:1.2–2.9), PHQ-9 (RR:1.8, 95%CI:0.8–3.8).

Limitations: Our study did not include clinical diagnosis of MDD by a specialized clinician, instead we used provisional diagnosis of probable MDD classified by the PHQ-9 algorithm as a reference standard in diagnostic performance evaluations

Conclusion: Depression screening tools varied in detection of postpartum MSD. The PHQ-2 would prompt fewer referrals and showed strong epidemiologic association with a cofactor.

Keywords

Postpartum depression; Screening; sub-Saharan Africa; CESD-10; EPDS; PHQ-9; PHQ-2

Introduction

Depression is a leading cause of maternal morbidity and mortality during pregnancy and the first year postpartum (Woody et al., 2017). It is the most common complication of the perinatal period (Stein et al., 2014), and disproportionately impacts women in low-income settings (Woody et al., 2017). Maternal depressive symptoms during the peripartum period may lead to a wide spectrum of adverse outcomes, including maternal suicide, adverse perinatal outcomes, and potentially child-adolescent behavioral and mental health problems (Stein et al., 2014).

Integration of depression screening into maternal and child health (MCH) services is increasingly encouraged by initiatives including the World Health Organization Mental Health Gap Action Programme (WHO mhGAP) (World Health Organization, 2016), the United Nations Sustainable Development Goals, and the Grand Challenges in Global Mental Health (Collins et al., 2011). These initiatives call for task-shifting to provide case identification and a range of interventions for mental disorders within routine care settings (Collins et al., 2011; Rahman et al., 2013). In sub-Saharan Africa (SSA), particularly in Kenya, MCH services are widely attended (Kenya Demographic and Health Survey, 2015)

and offer a high-impact access point for depression screening (Rahman et al., 2013). MCH clinics may be critical for maternal mental health screening as mental health issues have the potential to amplify other adverse MCH outcomes (Stein et al., 2014). However, currently MCH clinics in Kenya and elsewhere in SSA do not routinely offer depression screening and guidelines do not exist for perinatal depression screening in many low- and middle-income countries (LMICs). The Kenya Mental Health Policy (2015–2030) commits to increasing mental health service availability across the health system, including MCH, but does not name a preferred screening scale (Republic of Kenya Ministry of Health, 2015).

Multiple depression screening tools are available and utilized within perinatal populations (Gelaye et al., 2016; Tsai et al., 2013), yet it is unclear which are optimal for wide-scale deployment among pregnant and postpartum populations in MCH clinics in SSA. The Edinburgh Postnatal Depression Scale (EPDS) (Cox et al., 1987) is the most commonly used screening scale in perinatal populations in SSA, followed by the Patient Health Questionnaire-9 (PHQ-9) (Kroenke and Spitzer, 2002; Tsai et al., 2013) which is increasingly adopted for widescale depression screening in SSA. Other general depression screening instruments have been utilized in perinatal populations, including the Center for Epidemiologic Studies Depression Scale (CESD-10) (Andresen et al., n.d.; Dadi et al., 2020; Radloff, 1977) and the brief, 2-item version of the Patient Health Questionnaire (PHQ-2) (Kroenke et al., 2003a; Smith et al., 2010). Prior evaluations of depression screening tools have assessed diagnostic performance metrics (e.g., sensitivity, specificity) (Gelaye et al., 2016; Tsai et al., 2013). However, these varied depression screening tools have not been compared side-by-side for diagnostic performance, associations with known cofactors, and diagnostic yield.

Our study objective was to compare four tools (CESD-10, EPDS, PHQ-9, PHQ-2) commonly used for perinatal depression screening globally for a range of performance characteristics.

Methods

Study setting and population

This analysis was nested in the PrEP implementation for Mothers in Antenatal Care (PrIMA) study, a cluster randomized trial conducted within public sector MCH clinics in Kenya. The PrIMA RCT was designed to compare two models for pre-exposure prophylaxis implementation among pregnant women (NCT03070600) (Dettinger et al., 2019). Participants were enrolled from 20 MCH clinics within Homa Bay and Siaya counties of Western Kenya. Pregnant women were eligible for enrollment if they were HIV-negative, 15 years old, and were able to provide consent.

Data collection

Questionnaires were verbally administered by study nurses in Kiswahili, Dholuo, or English languages. This took place in private spaces to ensure confidentiality; participants were not given the option to self-complete depressive screening tools. Participants were surveyed about demographics, pregnancy history, partner characteristics, and psychosocial factors

including experience of depressive symptoms. All data collection instruments, including depression screening scales, were formally forward and backward translated from English into Kiswahili and Dhuluo. The data collection instrument was field tested by study staff and refined as needed. The depression screening scales included in this study were originally developed as self-assessments. In the present study, these scales were administered by trained study nurses to ensure consistency of administration across a study population of varied literacy and low familiarity with self-administered questionnaires. Our use of trained data collectors to gather depressive symptom data ensures comparability across similar studies, as most studies conducted to date that report depressive symptom data among similar populations also employ trained study staff to administer these scales.

Depression screening instruments

In this comparative analysis, we evaluated depressive symptoms at the 6-week postpartum visit using four depression screening scales: CESD-10, EPDS, PHQ-9, and PHQ-2. The CESD-10 and EPDS are 10-item screening scales where each item depicts a discrete depressive symptom and participants rate each item between 0 to 3 based on past-week frequency. Higher total scores indicate higher severity of depressive symptoms; scores range from 0–30. Symptoms of moderate-to-severe depression (MSD) was defined as CESD-10 score of ≥ 10 , the validated cut-point (Andresen et al., n.d.). We defined MSD symptoms with the EPDS as a score of ≥ 13 (Levis et al., 2020).

Similarly, the PHQ-9 is a 9-item screening scale assessing past two-week frequency of symptoms. Items are rated between 0–3 (range: 0–24); higher scores indicate greater severity. Scores of 10 or greater denote MSD symptoms (Kroenke et al., 2003b). The shorter PHQ-2 is comprised of two items: “Little interest or pleasure in doing things” and “Feeling down, depressed, or hopeless”. Scores range from 0–6. A cut-point of three or greater denotes likely major depression (Kroenke et al., 2003b). Collection of depressive symptom data did not begin at study activation for all scales, thus we do not have data in all scales for all participants.

Diagnostic algorithm for major depressive disorder

Information from the PHQ-9 can be utilized in a scoring algorithm to establish provisional diagnoses for probable major depressive disorder (MDD) based on the Diagnostic and Statistical Manual of Mental Disorders (DSM-V) (Kroenke and Spitzer, 2002; Spitzer et al., 1999). The original validation study of PHQ-9 diagnostic scoring found 85% overall accuracy, 75% sensitivity and 90% specificity compared to physician diagnosis with the DSM-III (Spitzer et al., 1999). An individual is considered to have probable MDD if they responded “more than half of the days” about their experience of at least five items in the past two weeks and if one these items was “little interest or pleasure in doing things” or “feeling down, depressed, or hopeless” (Kroenke et al., 2001). The item describing suicidal ideation contributes to the criteria at any frequency.

Defining predictors of perinatal depression

Experiencing violence inflicted by an intimate partner during pregnancy is well-established as a strong predictor of postpartum depression (Devries et al., 2013). About 2–14% of

women worldwide experience intimate partner violence (IPV) which increases risk of postpartum depression by at least two-fold (Devries et al., 2013). We examined the strength of association between IPV reported before 6 weeks postpartum and MSD symptoms in postpartum defined by each scale. We assessed IPV using the four-item “Hurt, Insult, Threaten, Scream” (HITS) scale with a cut-point of 10 to define IPV (absolute score range: 4–20) (Rabin et al., 2009). One’s ability to rely on a social network for emotional and logistical support is also associated with postpartum depression (Dadi et al., 2020). We used the 18-item Medical Outcomes Study Social Support Survey (MOS-SSS) to assess level of social support; participants identify how frequently they find various types of support when needed (higher scores denote higher social support, range: 18–90) (Sherbourne and Stewart, 1991). We defined low social support as scores below 72; these participants reported being unable to receive support at least “most of the time” for each scenario (Gold et al., 2013). The HITS and MOS-SSS tools are commonly utilized in research studies among populations in SSA (Abrefa-Gyan et al., 2015; Levy et al., 2016; Rabin et al., 2009)

Statistical analysis

We estimated prevalence of MSD symptoms at 6 weeks postpartum with the CESD-10, EPDS, PHQ-9, and PHQ-2 using descriptive statistics and estimated 95% confidence intervals from standard errors clustered by site. We evaluated differences between each pair of prevalence estimates using McNemar’s tests. Chance-adjusted agreement between scales for their identification of early postpartum MSD symptoms was estimated by Cohen’s kappa coefficients.

Sensitivity, specificity, and area under the curve (AUC) were assessed using receiver operating characteristic (ROC) curves comparing postpartum MSD symptoms for the CESD-10 and EPDS, to probable MDD. We did not compare the PHQ-9 or PHQ-2 screening scores to the algorithm for probable MDD since comparing summed items to a function of the same items may provide biased diagnostic performance estimates. We evaluated frequency of symptom endorsement for each depression scale among participants identified with MSD symptoms using that scale. To estimate the strength of epidemiologic relationships between IPV (experienced before 6 weeks postpartum), social support (during pregnancy) and postpartum MSD symptoms. We fit generalized linear regression models with binomial family, log link, clustered by facility. Estimates were adjusted for pre-specified confounders (age, marital status, household crowding, and social support or IPV). Analyses were conducted in Stata 15.

Ethical considerations

The study protocol, informed consent forms, and data collection tools were approved by the Kenyatta National Hospital-University of Nairobi Ethics Research Committee and University of Washington Human Subjects Review Committee. All participants provided written informed consent.

Role of funding sources

This work was supported by the National Institute of Allergy and Infectious Disease (R01 AI125498 to GJS), the Eunice Kennedy Shriver National Institute of Child Health and

Human Development (F31HD101149 to AL, R01HD100201 to JP and R01 HD094630 to GJS). All grants were externally peer reviewed for scientific quality. The funding agencies had no role in the writing of the manuscript or the decision to submit it for publication.

Results

Population characteristics

The 3,605 study participants who attended both the enrollment and six week postpartum visits and were included in this study. Median age was 24 years (interquartile range [IQR]: 21–29), the vast majority (86%, 3069) married or cohabiting with their partner, and only 8% (284) were currently in school (Table 1). About 15% (523) of the women were employed and 10% (371) experienced household crowding defined as over three people per room in their residence (Melki et al., 2004). Median social support score was 76 (IQR: 63–88); 36% (1262) women experienced low social support. One in ten women (10%, 368) reported IPV during pregnancy or early postpartum.

Prevalence of MSD symptoms and agreement between screening scales

Frequency of moderate-to-severe depressive symptoms detected in early postpartum varied by tool. The highest frequency was measured using the CESD-10 where 12.6% (391/3098, 95% confidence interval [CI]: 7.3%–20.8%) of participants had MSD symptoms. The EPDS identified 9.2% (326/3533, 95% CI: 5.9%–14.2%) of participants as having MSD symptoms. About 4.9% (176/3576, 95% CI: 2.6%–9.1%) of participants had MSD symptoms with the PHQ-2 and 3.2% had MSD symptoms using the PHQ-9 (74/2328, 95% CI: 1.6%–6.0%). The differences in frequency of MSD symptoms between each pair of scales were statistically significant ($p < 0.05$) (Figure 1).

Only 0.6% (13/2328, 95% CI: 0.2%–1.5%) of participants had MSD symptoms in every scale; 17.8% (643/3605, 95% CI: 11.4%–26.8%) had MSD symptoms in at least one scale. There was fair to moderate agreement between tools for classification of MSD symptoms (“Fair”: 0.2–0.4, “Moderate”: 0.4–0.6). Between the EPDS and PHQ-2, Cohen’s kappa coefficient was 0.220; between the PHQ-9 and EPDS kappa was 0.246, and between the PHQ-9 and CESD-10 it was 0.257. Kappa was 0.269 between the PHQ-2 and CESD-10, 0.369 between the CESD-10 and EPDS, and 0.543 between the PHQ-9 and PHQ-2.

Diagnostic performance compared to MDD

Among those with depressive symptom information in all four scales ($n=2328$), about 1.7% ($n=39$, 95% CI: 0.8%–3.4%) had probable MDD in early postpartum using the algorithm to estimate MDD (Figure 1). In area under the receiver operating characteristic curve (AUROC) comparisons against probable MDD, the CESD-10 had an AUROC of 0.82; sensitivity 50%, specificity 91% at a score of 10. The EPDS had an AUROC of 0.75; 42% sensitivity, 92% specificity at a score of 13. Both scales demonstrated satisfactory diagnostic performance compared to probable MDD classification with the PHQ-9 algorithm (Figure 2).

Symptomatology

In our assessment of symptom endorsement among women with MSD symptoms defined by each tool, we found that the three longer scales (CESD-10, EPDS, PHQ-9) identified similar symptoms of feeling overwhelmed (“Everything was an effort” [92%], “Things were getting on top of me” [98%]) or having low energy (“feeling tired or having little energy” [96%]) as the most common among postpartum Kenyan women. A related symptom of anhedonia, or inability to feel pleasure, was the most common symptom reported by those with MSD symptoms defined by the PHQ-2 (92%). The remaining symptoms ranked differently by scale (Figure 3).

Detection of associations with known predictor

We found intimate partner violence experienced prior to 6 weeks postpartum was associated with MSD symptoms in early postpartum using each scale. The strength of association varied between scales when symptoms of MSD was defined using the screening thresholds recommended by tool developers. The strongest association was detected by the EPDS where those reporting IPV were 2.5-times more likely to experience MSD symptoms than those not reporting IPV. With the PHQ-2, risk for MSD symptoms among those exposed to IPV was 2.3-times higher than those not exposed. Experiencing IPV put women at 1.9-times higher risk for MSD symptoms in early postpartum with the CESD-10 and 1.8-times higher risk with the PHQ-9 compared to those without IPV. Relationships were robust to adjustment for pre-specified confounders.

Low social support during pregnancy also put women at twice the risk of MSD symptoms in early postpartum as measured with the CESD-10 and 1.5-times higher risk of MSD symptoms measured with the PHQ-2 (Figure 4). This relationship was not significant with MSD symptoms in the PHQ-9 or EPDS. Statistical significance was maintained after adjustment for confounders with MSD symptoms in the CESD10 ($p=0.002$), yet not the PHQ-2 ($p=0.116$). When we evaluated social support as a continuous score, we found a 1-unit increase in social support was statistically significantly associated with lower risk of MSD in the CESD-10 (RR:0.97, 95% CI:0.97–0.99, $p<0.001$), which was robust to confounding adjustment (aRR: 0.98, 95% CI: 0.97–0.99, $p<0.001$). We saw a trend toward significance when MSD was defined by the PHQ-2 (RR:0.98, 95% CI:0.97–1.00, $p=0.101$), yet not the PHQ-9 (RR:0.99, 95% CI:0.98–1.01, $p=0.765$) or the EPDS (0.99, 95% CI:0.97–1.01, $p=0.184$).

Sensitivity analyses

In sensitivity analyses, we repeated all assessments using higher severity-level cut-points. We used cut-points of EPDS 15, PHQ-9 15, and CESD-10 15 and did not repeat analyses for PHQ-2 since higher severity cut-points are not standard. Results from all analyses were consistent with our main findings. Prevalence patterns and kappa agreement did not appreciably change. Sensitivity decreased (40%) and specificity increased (100%) for both the CESD-10 and EPDS in ROC curve analyses at the revised cut-points. Symptomatology did not meaningfully change; the magnitude of effects of IPV and social support on postpartum depressive symptoms was higher, with wider 95% confidence intervals which encompassed original estimates of MSD symptoms. We repeated all analyses among the

subset of participants with depressive symptom data in all scales (n=2328). Results for MSD prevalence, diagnostic performance, and symptomatology did not change meaningfully. Cofactor associations were no longer statistically significant, likely due to insufficient power.

Discussion

Main findings

In this comparative evaluation of four commonly used depression screening scales in a large cohort of postpartum Kenyan women, we found that the CESD-10, EPDS, PHQ-9, and PHQ-2 varied widely in detection of moderate-to-severe depressive symptoms. Only about 2% of Kenyan women had MSD detected by all four scales. Depressive symptoms reported most frequently by women with MSD differed, indicating that scales identified women with different manifestations of depressive symptoms, potentially explaining the lack of strong agreement between scales. Our evaluation of the known epidemiologic relationships between IPV, social support, and postpartum MSD symptoms showed significant associations with moderate differences in strength of association between the scales. The CESD-10 and EPDS demonstrated high diagnostic performance to identify probable MDD, however these results should be interpreted with caution due to our use of a non-ideal reference standard. To our knowledge, this is the first study to compare the prevalence of MSD between the CESD-10, EPDS, PHQ-9, and PHQ-2 in a large cohort of postpartum African women. Our results highlight tradeoffs for tool selection which has implications for research and practice.

Interpretation

Prevalence of MSD symptoms in early postpartum ranged from 3%–13% depending on the depression screening scale used. As such, agreement between every pair of scales based on Cohen's kappa was also only slight-to-moderate. Our prevalence estimates are consistent with prevalence estimates in similar research and clinical settings of SSA (Dadi et al., 2020; Tsai et al., 2013). A recent systematic review and meta-analysis of postnatal depression in Africa revealed a pooled prevalence of 16.8% and yielded prevalence estimates similar to those in our study with specific depression screening tools: 21.4% with CESD-10 in Zimbabwe, 9.2% with EPDS in Sudan, and 7.0% with PHQ-9 in Ghana (Dadi et al., 2020). To date, most studies have estimated prevalence of MSD symptoms using a single screening scale. Our study was unique in concurrently assessing 4 tools and demonstrated up to a 2-fold difference in prevalence of depression between scales.

Our findings suggest that variability in depression prevalence observed between studies may be a result of different scales used in assessing depression (Dadi et al., 2020; Tsai et al., 2013). These differences have implications for who receives more-intensive (e.g., referral to specialized psychiatric care) versus less-intensive (e.g., psychological support, self-help messages) mental health intervention within health systems. The PHQ-9 would result in the lowest number of referrals for higher-intensity services (32 per 1000 women screened), referrals with the PHQ-2 would also be low (49 per 1000 women screened), while the CESD-10 scale would produce the highest volume (126 referrals per 1000

women screened). Multi-stage screening approaches that utilize brief and longer depression screening scales, may help optimize allocation of available resources for maternal mental health in low-resource settings (World Health Organization, 2016). The WHO recommends task-shifted depression evaluation by non-specialized healthcare providers through two screening questions consistent with the PHQ-2, followed by more comprehensive questions similar to the PHQ-9 (mhGAP approach) (World Health Organization, 2016). Our results support this approach in MCH settings.

We found slight-to-moderate agreement between scales. A prior study that evaluated agreement between the EPDS and PHQ-9 depression scales among 1500 pregnant women in Peru similarly found moderate agreement ($\kappa=0.36$). We found higher agreement between PHQ-2 and PHQ-9 than a prior study ($\kappa=0.17$) comparing PHQ-2 and PHQ-9 in 218 American pregnant women (Smith et al., 2010). Differences in agreement between scales illustrate the varied domains assessed by each scale, which would result in differences in *numbers* of women referred, but also in *which* women would be referred.

The CESD-10 and EPDS both demonstrated strong diagnostic performance with high sensitivity, specificity, and AUROC for MSD symptoms compared to probable MDD. Of note, the reference standard for probable MDD used in this analysis was the PHQ-9 scoring algorithm for DSM-V classification of probable MDD (Kroenke and Spitzer, 2002). The PHQ-9 algorithm approach was originally developed as a rapid, self-report version of the *Primary Care Evaluation of Mental Disorders* (PRIME-MD) that could be used to efficiently classify probable MDD (Kroenke and Spitzer, 2002). It has also been a reference standard in diagnostic validation studies for depression screening tools when physician-led clinical diagnostic interviews were not available (Corson et al., 2004; Kroenke et al., 2009). Since the PHQ-9 and PHQ-2 screening scores are comprised of the same items used to define the MDD algorithm, we did not perform ROC analyses comparing the PHQ-9 and PHQ-2 to the MDD reference standard. Overall, the CESD-10 demonstrated high sensitivity, specificity, and AUROC (0.82), while the EPDS had lower sensitivity yet was very accurate in distinguishing non-cases (AUROC: 0.75), yet these results should be interpreted cautiously due to our use of a non-ideal reference standard. A more formal diagnostic performance comparison which utilizes a diagnostic interview as the reference standard is needed to more rigorously evaluate the performance of these scales among perinatal African women.

All tools have been separately validated in at least one study in either African perinatal (EPDS (Chibanda et al., 2010), PHQ-9 (Weobong et al., 2009)) or general populations (CESD-10 (Baron et al., 2017), PHQ-2 (Gelaye et al., 2016)). To ensure those who would benefit from more intensive psychosocial treatment are not missed due to a false negative result, high sensitivity is important. In parallel, optimal specificity is necessary to avoid misallocation of higher-intensity, specialized provider time to individuals with a false positive screening result (Thombs and Ziegelstein, 2013). In clinical settings which rely on task-shifted mental health screening and service provision, the ideal balance between sensitivity and specificity may be influenced by available interventions, costs, and accessibility (Agency for Healthcare Research and Quality, 2013). Accepting a loss in sensitivity to ensure high specificity may help optimize resource allocation (Agency for Healthcare Research and Quality, 2013).

We found significant associations between MSD symptoms and IPV using all four screening tools, consistent with the published literature (Devries et al., 2013). The association was the strongest with EPDS (RR: 2.5) and lowest with PHQ-9 (relative risk: 1.8). We also saw differences in effect size and statistical significance for relationships between MSD symptoms and social support. Our observation of varied strength of association has implications for interpreting epidemiologic studies that assess cofactors of MSD symptoms using different screening tools. The variation in prevalence estimates for MSD symptoms across the scales contributes to differences in relationships with cofactors, thus differences in such relationships between studies should not be over-interpreted.

Strengths

Our data was collected in a large ($n > 3000$), multisite randomized cluster trial and included multiple commonly used depressive symptom scales validated in African settings. Participants were followed over time, establishing temporality of relationships between cofactors and depression. This is the first study to directly compare the performance of the CESD-10, EPDS, PHQ-9, and PHQ-2 in a large cohort of postpartum African women.

Limitations

The PrIMA RCT did not enroll women living with HIV based on the main aim of evaluating HIV prevention strategies, which limits generalizability of our findings in settings with high HIV prevalence. Our study did not include clinical diagnosis of MDD by a specialized clinician, instead we used provisional diagnosis of probable MDD classified by the PHQ-9 algorithm as a reference standard in diagnostic performance evaluations (Kroenke and Spitzer, 2002). This method allows use of PHQ-9 items for provisional MDD classification in the absence of clinical evaluation and has been used previously to evaluate diagnostic performance of depression screening tools (Corson et al., 2004; Kroenke et al., 2009).

Conclusion

Depression screening tools varied in detection of moderate-to-severe depressive symptoms in early postpartum between four depression screening instruments used among perinatal populations worldwide: CESD-10, EPDS, PHQ-9, and PHQ-2. These depression tools had moderate differences in symptomatology and strength of association with intimate partner violence and social support. We also saw moderate differences in diagnostic performance between the CESD-10 and EPDS compared to PHQ-9 algorithm for MDD. All scales performed well, and the PHQ-2 emerged as the depression screening tool that would balance low referral burden to the health system, while maintaining valid estimates of known epidemiologic relationships. Additionally, the brevity of the PHQ-2 as a two-item scale has advantages for application in maternal child health clinics. These results support the WHO recommendation for multi-stage screening with PHQ-2 followed by administration of a longer depression screening scale. Overall, our findings have implications for tool selection in research and clinical practice addressing depression among pregnant and postpartum women in sub-Saharan African settings.

Acknowledgments

We thank the study participants for their time and information.

Role of funding sources

This work was supported by the National Institute of Allergy and Infectious Disease (R01 AI125498 to GJS), the Eunice Kennedy Shriver National Institute of Child Health and Human Development (F31HD101149 to AL, R01HD100201 to JP and R01 HD094630 to GJS). All grants were externally peer reviewed for scientific quality. The funding agencies had no role in the writing of the manuscript or the decision to submit it for publication.

Disclosure of interests:

This paper represents the opinions of the authors and is not meant to represent the position or opinions of organizations, nor the official position of any staff members. Ms. Larsen reports grants from NIH during the conduct of the study. Dr. Bhat reports grants from NIH, grants from PCORI, and grants from Perigee Foundation during the conduct of the study. Dr. John-Stewart reports grants from NIH, grants from CDC, grants from Thrasher, personal fees from UpToDate, personal fees from UW, grants from IMPAACT, outside the submitted work. Dr. Kinuthia reports grants from NIH, during the conduct of the study. Dr. Pintye reports grants from NIH during the conduct of the study.

References

- Abrefa-Gyan T, Wu L, Lewis MW, 2015. Social support and support groups among people with HIV/AIDS in Ghana. *10.1080/00981389.2015.1084969* 55, 144–160. *10.1080/00981389.2015.1084969*
- Agency for Healthcare Research and Quality, 2013. Efficacy and Safety of Screening for Postpartum depression. Report 106.
- Andresen EM, Malmgren JA, Carter WB, Patrick DL, n.d. Screening for depression in well older adults: evaluation of a short form of the CES-D (Center for Epidemiologic Studies Depression Scale). *Am. J. Prev. Med* 10, 77–84. [PubMed: 8037935]
- Andresen EM, Malmgren JA, Carter WB, Patrick DL, n.d. Screening for depression in well older adults: evaluation of a short form of the CES-D (Center for Epidemiologic Studies Depression Scale). *Am. J. Prev. Med* 10, 77–84. [PubMed: 8037935]
- Baron EC, Davies T, Lund C, 2017. Validation of the 10-item Centre for Epidemiological Studies Depression Scale (CES-D-10) in Zulu, Xhosa and Afrikaans populations in South Africa. *BMC Psychiatry* 17, 6. *10.1186/s12888-016-1178-x* [PubMed: 28068955]
- Chibanda D, Mangezi W, Tshimanga M, Woelk G, Rusakaniko P, Stranix-Chibanda L, Midzi S, Maldonado Y, Shetty AK, 2010. Validation of the Edinburgh Postnatal Depression Scale among women in a high HIV prevalence area in urban Zimbabwe. *Arch. Womens. Ment. Health* 13, 201–206. *10.1007/s00737-009-0073-6* [PubMed: 19760051]
- Collins PY, Patel V, Joestl SS, March D, Insel TR, Daar AS, Scientific Advisory Board and the Executive Committee of the Grand Challenges on Global Mental Health, Anderson W, Dhansay MA, Phillips A, Shurin S, Walport M, Ewart W, Savill SJ, Bordin IA, Costello EJ, Durkin M, Fairburn C, Glass RI, Hall W, Huang Y, Hyman SE, Jamison K, Kaaya S, Kapur S, Kleinman A, Ogunniyi A, Otero-Ojeda A, Poo M-M, Ravindranath V, Sahakian BJ, Saxena S, Singer PA, Stein DJ, 2011. Grand challenges in global mental health. *Nature* 475, 27–30. *10.1038/475027a* [PubMed: 21734685]
- Corson K, Gerrity M, Dobscha S, 2004. Screening for depression and suicidality in a VA primary care setting: 2 items are better than 1 item - PubMed. *Am. J. Manag. Care* 11, 839–845.
- Cox JL, Holden JM, Sagovsky R, 1987. Detection of Postnatal Depression. *Br. J. Psychiatry* 150, 782–786. *10.1192/bjp.150.6.782* [PubMed: 3651732]
- Dadi AF, Akalu TY, Baraki AG, Wolde HF, 2020. Epidemiology of postnatal depression and its associated factors in Africa: A systematic review and meta-analysis. *PLoS One*. *10.1371/journal.pone.0231940*
- Dettinger JC, Kinuthia J, Pintye J, Mwongeli N, Gómez L, Richardson BA, Barnabas R, Wagner AD, O'Malley G, Baeten JM, John-Stewart G, 2019. PrEP Implementation for Mothers in

- Antenatal Care (PrIMA): Study protocol of a cluster randomised trial. *BMJ Open* 9. 10.1136/bmjopen-2018-025122
- Devries KM, Mak JY, Bacchus LJ, Child JC, Falder G, Petzold M, Astbury J, Watts CH, 2013. Intimate Partner Violence and Incident Depressive Symptoms and Suicide Attempts: A Systematic Review of Longitudinal Studies. *PLoS Med.* 10, e1001439. 10.1371/journal.pmed.1001439 [PubMed: 23671407]
- Gelaye B, Rondon M, Araya R, Williams M, 2016. Epidemiology of maternal depression, risk factors, and child outcomes in low-income and middle-income countries. *Lancet Psychiatry* 3, 973–982. [PubMed: 27650773]
- Gold KJ, Spangenberg K, Wobil P, Schwenk TL, 2013. Depression and risk factors for depression among mothers of sick infants in Kumasi, Ghana. *Int. J. Gynecol. Obstet* 120, 228–231. 10.1016/j.ijgo.2012.09.016
- Kenya Demographic and Health Survey, 2015.
- Kroenke K, Spitzer RL, 2002. The PHQ-9: A new depression diagnostic and severity measure. *Psychiatr. Ann* 10.3928/0048-5713-2002090106
- Kroenke K, Spitzer RL, Williams JBW, 2003a. The patient health questionnaire2: Validity of a two-item depression screener. *Med. Care* 41, 1284–1292. 10.1097/01.MLR.0000093487.78664.3C [PubMed: 14583691]
- Kroenke K, Spitzer RL, Williams JBW, 2003b. The patient health questionnaire2: Validity of a two-item depression screener. *Med. Care* 41, 1284–1292. 10.1097/01.MLR.0000093487.78664.3C [PubMed: 14583691]
- Kroenke K, Spitzer RL, Williams JBW, 2001. The PHQ-9: Validity of a brief depression severity measure. *J. Gen. Intern. Med* 16, 606–613. 10.1046/j.1525-1497.2001.016009606.x [PubMed: 11556941]
- Kroenke K, Strine TW, Spitzer RL, Williams JBW, Berry JT, Mokdad AH, 2009. The PHQ-8 as a measure of current depression in the general population. *J. Affect. Disord* 114, 163–173. 10.1016/j.jad.2008.06.026 [PubMed: 18752852]
- Levis B, Negeri Z, Sun Y, Benedetti A, Thombs BD, 2020. Accuracy of the Edinburgh Postnatal Depression Scale (EPDS) for screening to detect major depression among pregnant and postpartum women: Systematic review and meta-analysis of individual participant data. *BMJ*. 10.1136/bmj.m4022
- Levy ME, Ong'wen P, Lyon ME, Cohen CR, D'Angelo LJ, Kwena Z, Wolf HT, 2016. Low Social Support and HIV-Related Stigma Are Highly Correlated among Adolescents Living With HIV in Western Kenya. *J. Adolesc. Heal* 58, S82. 10.1016/J.JADOHEALTH.2015.10.177
- Melki IS, Beydoun HA, Khogali M, Tamim H, Yunis KA, National Collaborative Perinatal Neonatal Network (NCPNN), 2004. Household crowding index: a correlate of socioeconomic status and inter-pregnancy spacing in an urban setting. *J. Epidemiol. Community Heal* 58, 476–480. 10.1136/jech.2003.012690
- Rabin RF, Jennings JM, Campbell JC, Bair-Merritt MH, 2009. Intimate partner violence screening tools: a systematic review. *Am. J. Prev. Med* 36, 439–445.e4. 10.1016/j.amepre.2009.01.024 [PubMed: 19362697]
- Radloff LS, 1977. The CES-D Scale. *Appl. Psychol. Meas* 1, 385–401. 10.1177/014662167700100306
- Rahman A, Surkan PJ, Cayetano CE, Rwagatare P, Dickson KE, 2013. Grand Challenges: Integrating Maternal Mental Health into Maternal and Child Health Programmes. *PLoS Med.* 10, e1001442. 10.1371/journal.pmed.1001442 [PubMed: 23667345]
- Republic of Kenya Ministry of Health, 2015. Kenya Mental Health Policy 2015 to 2030. Nairobi, Kenya.
- Sherbourne CD, Stewart AL, 1991. The MOS social support survey. *Soc. Sci. Med* 32, 705–714. 10.1016/0277-9536(91)90150-B [PubMed: 2035047]
- Smith MV, Gotman N, Lin H, Yonkers KA, 2010. Do the PHQ-8 and the PHQ-2 accurately screen for depressive disorders in a sample of pregnant women? *Gen. Hosp. Psychiatry* 32, 544–8. 10.1016/j.genhosppsy.2010.04.011 [PubMed: 20851275]

- Spitzer RL, Kroenke K, Williams JBW, 1999. Validation and utility of a self-report version of PRIME-MD: The PHQ Primary Care Study. *J. Am. Med. Assoc* 282, 1737–1744. 10.1001/jama.282.18.1737
- Stein A, Pearson RM, Goodman SH, Rapa E, Rahman A, McCallum M, Howard LM, Pariante CM, 2014. Effects of perinatal mental disorders on the fetus and child. *Lancet (London, England)* 384, 1800–19. 10.1016/S0140-6736(14)61277-0 [PubMed: 25455250]
- Thombs BD, Ziegelstein RC, 2013. Depression screening in primary care: Why the Canadian Task Force on Preventive Health Care did the right thing. *Can. J. Psychiatry* 10.1177/070674371305801207
- Tsai AC, Scott JA, Hung KJ, Zhu JQ, Matthews LT, Psaros C, Tomlinson M, 2013. Reliability and Validity of Instruments for Assessing Perinatal Depression in African Settings: Systematic Review and Meta-Analysis. *PLoS One* 8, e82521. 10.1371/journal.pone.0082521 [PubMed: 24340036]
- Weobong B, Akpalu B, Doku V, Owusu-Agyei S, Hurt L, Kirkwood B, Prince M, 2009. The comparative validity of screening scales for postnatal common mental disorder in Kintampo, Ghana. *J. Affect. Disord* 113, 109–117. 10.1016/j.jad.2008.05.009 [PubMed: 18614241]
- Woody CA, Ferrari AJ, Siskind DJ, Whiteford HA, Harris MG, 2017. A systematic review and meta-regression of the prevalence and incidence of perinatal depression. *J. Affect. Disord* 219, 86–92. 10.1016/j.jad.2017.05.003 [PubMed: 28531848]
- World Health Organization, 2016. mhGAP Intervention Guide for mental, neurological, and substance use disorders in non-specialized health settings.

Highlights

- We compared four depression scales (EPDS, PHQ-9/-2, CESD-10) in postpartum women
- Scales had differences in diagnostic yield and epidemiologic associations
- Experts speculate perinatal depression estimates differ across studies due to tools used
- We confirm this with empiric data about tool influence on depression estimates
- PHQ-2 may provide optimal features for wide-scale use in maternal child health

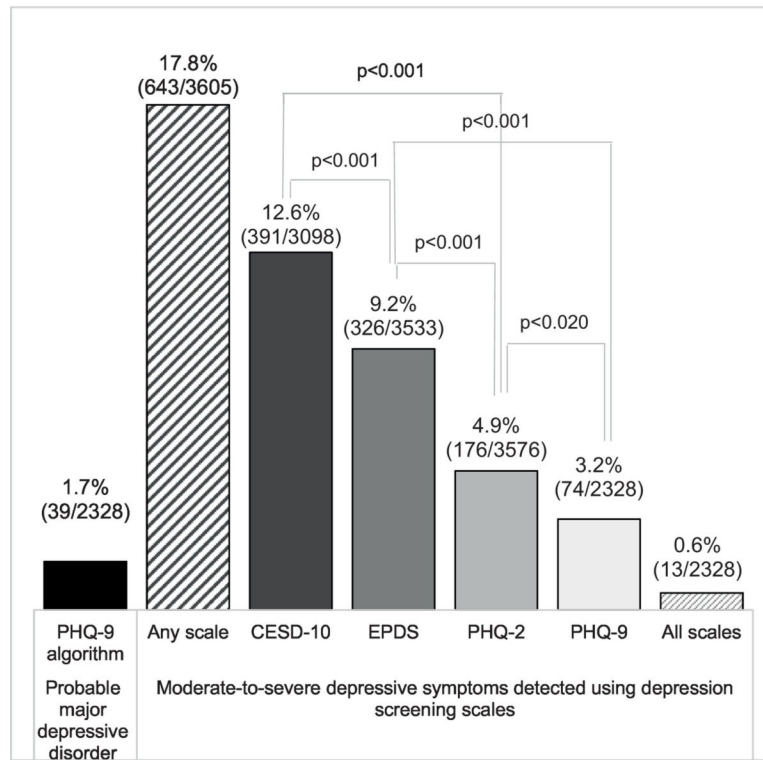
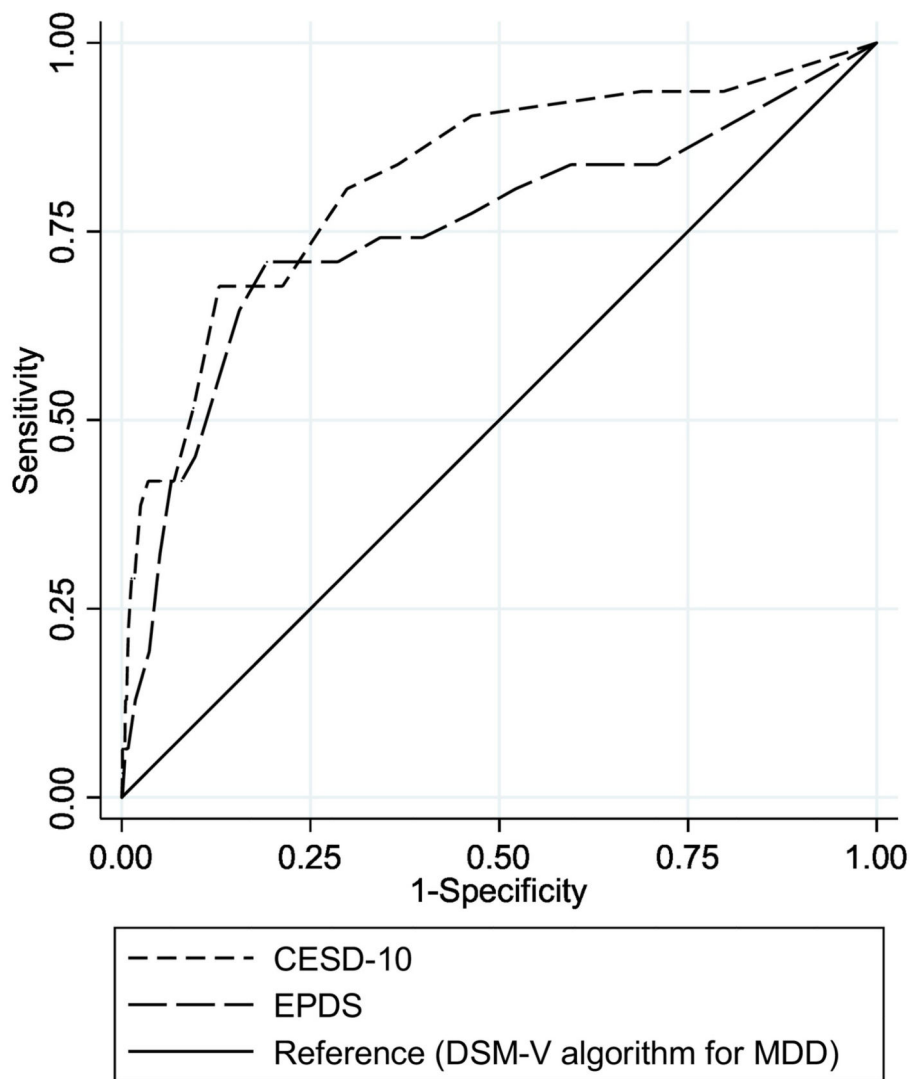


Figure 1. Comparing frequency of moderate-to-severe depressive symptoms using four depression screening scales among postpartum Kenyan women



Depression symptom scale	Sensitivity	Specificity	AUROC
CESD-10 score ≥ 10	50.0 %	90.6 %	0.822
EPDS score ≥ 13	41.7 %	92.1 %	0.752

Figure 2.
Comparing diagnostic performance of the CESD-10 and EPDS depression scales among postpartum Kenyan women (n=2328)

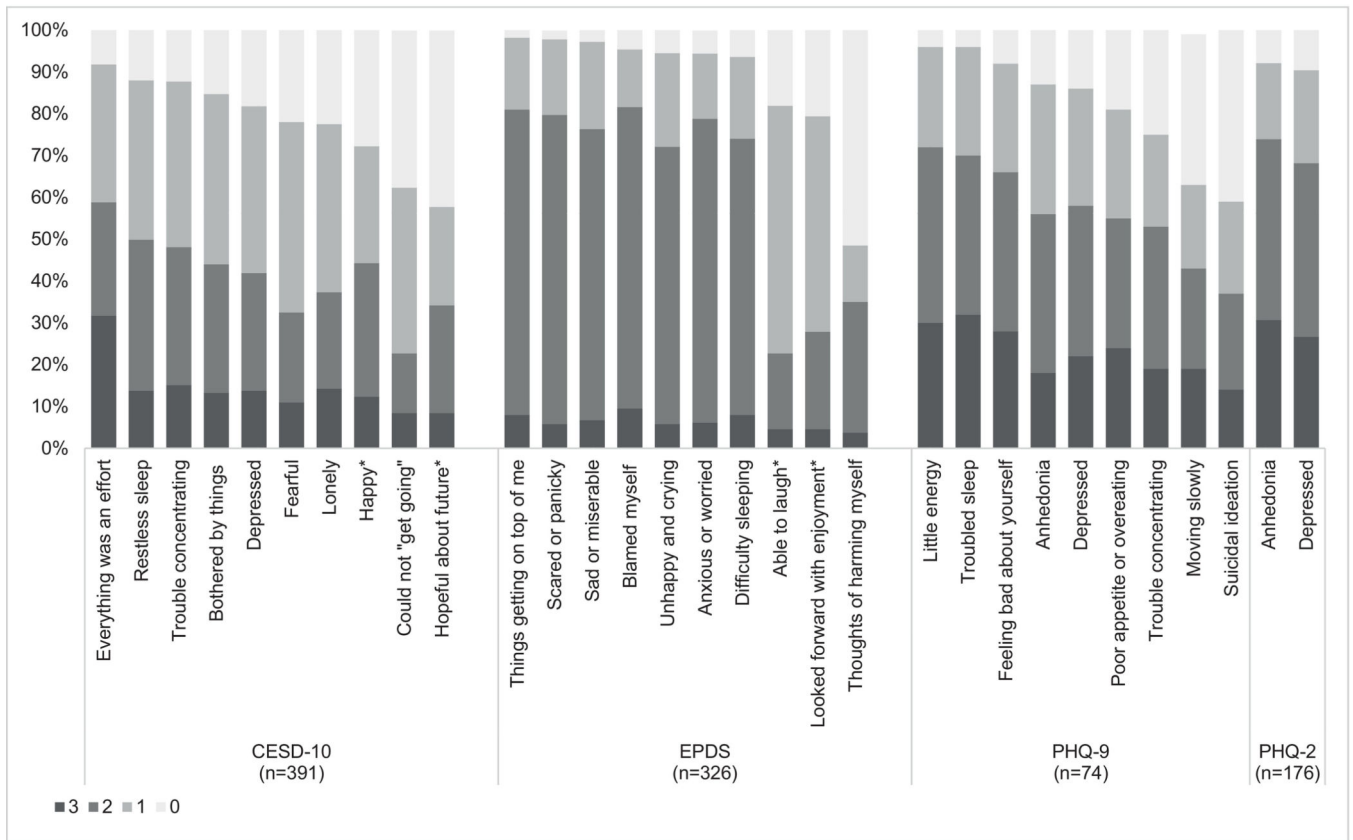


Figure 3. Symptom endorsement for four depression scales among postpartum Kenyan women with MSD

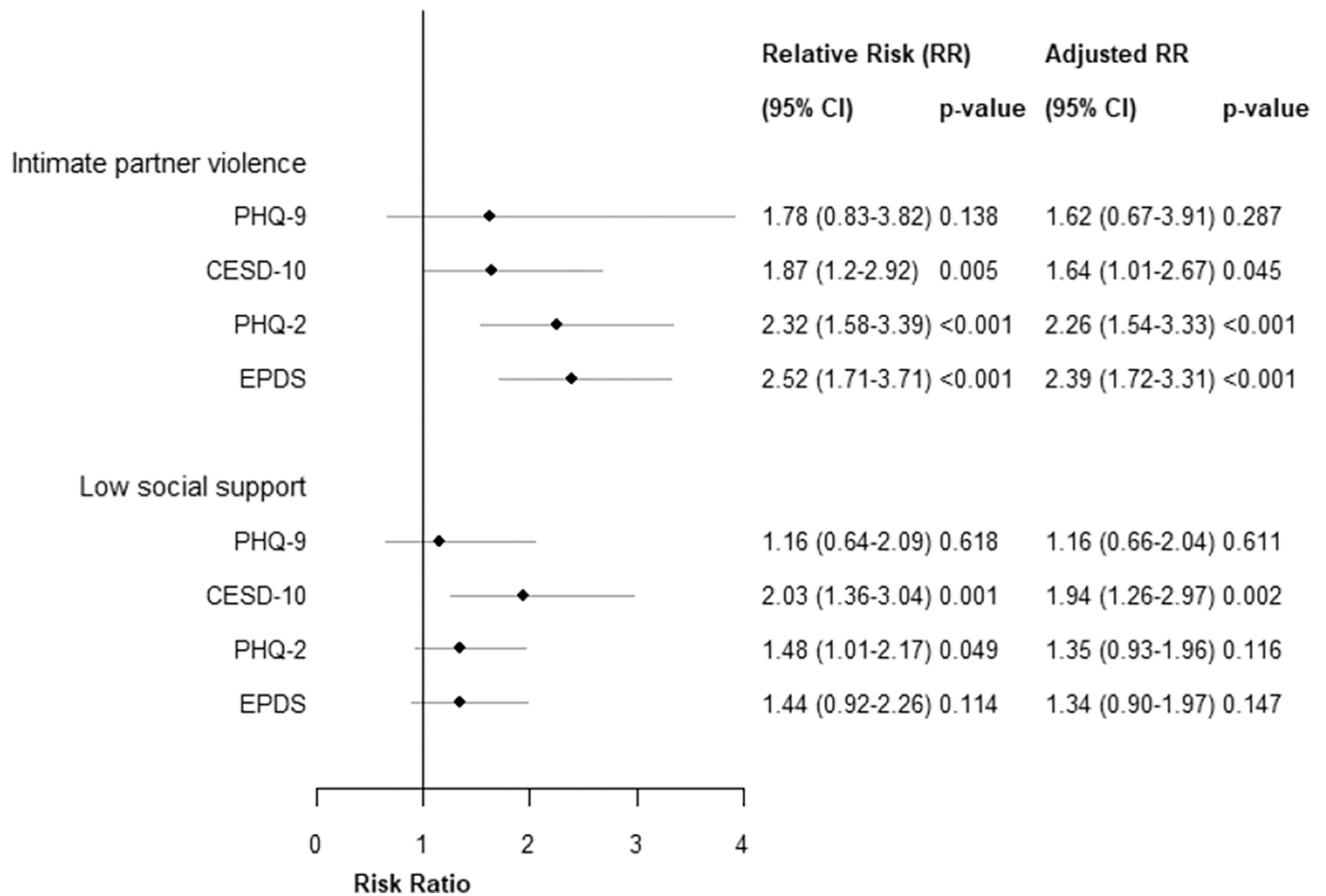


Figure 4. Epidemiologic associations between intimate partner violence and low social support with MSD symptoms among postpartum Kenyan women

Table 1.

Characteristics of PrIMA participants followed through six weeks postpartum

Characteristic	N	Overall (n=3605)	
		N or median	% or IQR
Demographic characteristics			
Age (years)	3603	24	21–29
Adolescents and young adults (<25 years)	3603	1991	55.3%
Married/living with a partner	3562	3069	86.2%
Currently in school	3552	284	8.0%
Completed education (years)	3524	10	8–12
Regularly employed	3552	523	14.7%
Household crowding (>3 people/room)	3567	371	10.4%
Pregnancy history			
Gestational age at enrollment (weeks)	3585	24	20–29
Ever pregnant before	3588	2750	76.6%
Number of pregnancies	3593	2	2–3
Partnership and sexual behavior characteristics			
Partner age difference >10 years *	2772	450	16.2%
Partner HIV-positive *	3318	172	5.2%
Partner HIV status unknown	3605	963	26.7%
Self-perceived HIV risk (within next year) “Extremely/very likely”	3583	1701	47.5%
Transactional sex ever in last 6 mo.	3587	59	1.6%
Forced to have sex against her will ever in last 6 mo.	3587	191	5.3%
Psychosocial characteristics			
Ever drink alcohol	3574	140	3.9%
Social support score ^a	3517	76	63–88
Low social support (MOS-SSS score <72)	3517	1262	35.9%
Intimate partner violence ^b (HITS score ≥ 10)	3578	368	10.3%

* Among those with a current partner

^aWe evaluated social support using the 18-item Medical Outcomes Study social support score (MOS-SSS), defining low social support as scores below 72 (Low social support: MOS-SSS score <72 = “Yes”, MOS-SSS score ≥ 72 = “No”)

^bWe evaluated intimate partner violence using the 4-item Hurt, Insult, Threaten, and Scream scale (HITS), defining intimate partner violence as scores of 10 and above (IPV: HITS score ≥ 10 = “Yes”, HITS score <10 = “No”)