Arch Womens Ment Health (2016) 19:883–890
DOI 10.1007/s00737-016-0632-6

ORIGINAL ARTICLE

The experience of depression, anxiety, and mania among perinatal women

J. Jo Kim[1,2] · Richard K. Silver[1,2] · Rita Elue[1] · Marci G. Adams[1] · Laura M. La Porte[1] · Li Cai[3] · Jong Bae Kim[2] · Robert D. Gibbons[2]

Received: 8 March 2016 / Accepted: 30 April 2016 / Published online: 17 May 2016

Abstract We assessed differential item functioning (DIF) based on computerized adaptive testing (CAT) to examine how perinatal mood disorders differ from adult psychiatric disorders. The CAT-Mental Health (CAT-MH) was administered to 1614 adult psychiatric outpatients and 419 perinatal women with IRB approval. We examined individual item-level differences using logistic regression and overall score differences by scoring the perinatal data using the original bifactor model calibration based on the psychiatric sample data and a new bifactor model calibration based on the perinatal data and computing their correlation. To examine convergent validity, we computed correlations of the CAT-MH with contemporaneously administered Edinburgh Postnatal Depression Scales (EPDS). The rate of major depression in the perinatal sample was 13 %. Rates of anxiety, mania, and suicide risk were 5, 6, and 0.4 %, respectively. One of 66 depression items, one of 69 anxiety items, and 15 of 53 mania items exhibited DIF (i.e., failure to discriminate between high and low levels of the disorder) in the perinatal sample based on the psychiatric sample calibration. Removal of these items resulted in correlations of the original and perinatal calibrations of $r = 0.983$ for depression, $r = 0.986$ for anxiety, and $r = 0.932$ for mania. The 91.3 % of cases were concordantly categorized as either "at-risk" or "low-risk" between the EPDS and the perinatal calibration of the CAT-MH. There was little evidence of DIF for depression and anxiety symptoms in perinatal women. This was not true for mania. Now calibrated for perinatal women, the CAT-MH can be evaluated for longitudinal symptom monitoring.

Keywords Computerized adaptive testing · Item response theory · Perinatal depression · Perinatal mania

Introduction

Perinatal depression affects 10–20 % of pregnant and postpartum women. If untreated during pregnancy, women who are depressed are more likely to deliver preterm (Dayan et al. 2006), have a low birth weight or growth-restricted infant (Rahman et al. 2004), or develop preeclampsia (Kurki et al. 2000). Untreated depressed mothers may breastfeed less often and, if they do, are less likely to breastfeed exclusively (Hatton et al. 2005). Their parenting capability may be impaired over and above other risk factors, and adversely affect their children's cognitive, emotional, and behavioral development (Van Doesum et al. 2007; Cogill et al. 1986; Whitaker et al. 2006).

Studies show that screening can aid in detection of perinatal depression (Evins et al. 2000; Heneghan et al. 2000; Georgiopoulos et al. 2001), and universal screening is widely recommended (Siu et al. 2016; Earls 2010; Hirst and Moutlier 2010; ACOG 2015). Unfortunately, a common practice in mental health screening is to take a measurement tool developed in patients with one indication and then use it in the assessment of patients with another indication or with comorbid disorders that may impact the validity of the scale. Differences in the parameters of the score distributions between the two indications (e.g., mean and variance) are then interpreted as if they represent differences in the underlying disease or construct of interest. This assumes that the properties of the administered items are invariant between the two populations of interest. However, an

✉ J. Jo Kim
jkim@northshore.org

[1] NorthShore University HealthSystem, 2650 Ridge Avenue, Walgreen Suite 1507, Evanston, IL 60201, USA

[2] University of Chicago, Chicago, USA

[3] University of California Los Angeles, Los Angeles, USA

item that is an excellent discriminator between high and low levels of depression in a psychiatric clinic may not be in a primary care setting, emergency department, oncology center, or perinatal clinic. Inclusion of this item may provide a biased estimate of the underlying latent dimension of interest (e.g., depressive severity).

The importance of measurement-based assessment and referral for depression is recognized as a necessary first step in improving care (deGruy and Pincus 1996). However, in busy obstetrical practices, it is generally unfeasible to conduct psychometrically sound psychiatric evaluations. Any strategy that reduces the burden of empirically based assessment has the potential to reorient the care system toward improved treatment and improved outcomes, without which the value of screening alone is likely limited (ACOG 2015). Computerized adaptive testing (CAT) permits a clinician to gather important clinical information via self-report in a way that dramatically reduces the burden on both the patient and the caregiver while maximizing the precision of measurement (Gibbons 2012, 2013, 2014). The basic idea is to administer an item, compute an estimate of severity (e.g., of depression) and the uncertainty in that estimate, and then select the next most informative item remaining in the bank of items. This process continues until the uncertainty falls below a previously defined threshold (e.g., 5 points on a 100-point scale). CAT depends on a previous calibration of the items based on Item Response Theory (IRT). The paradigm shift is from traditional measurement based on classical test theory, which fixes the number of items administered and allows measurement uncertainty to vary as is the case for the Edinburgh Postnatal Depression Scale (EPDS) (Cox et al. 1987), to IRT-based CAT, which fixes measurement uncertainty and allows the items to vary. The result is both a reduction in the number of items needed to measure mental health constructs such as depression and increased precision of measurement. Unlike traditional measurement, CAT administers a small number of items that are targeted to a patient's specific impairment level (Weiss 1985).

Traditional educational testing examples of CAT have been based on unidimensional IRT since they have been applied to essentially unidimensional ability domains such as mathematical aptitude. However, mental health constructs are inherently multidimensional, and therefore, CAT based on multidimensional IRT models, e.g., the bifactor model (Gibbons and Hedeker 1992), is required (Gibbons 2014). The net result is accurate measurement and monitoring of mental health constructs such as depressive severity and screening diagnoses such as major depressive disorder (MDD) in a few minutes, either in or outside of the clinic, via the internet. When based on large item banks (e.g., the CAT-Mental Health (CAT-MH) is based on 1008 items overall), the same items are not repeatedly administered during subsequent evaluations, so response set bias associated with traditional fixed-length measures, e.g., EPDS, Patient Health Questionnaire-9 (Kroenke et al. 2001), Hamilton Rating Scale for Depression (Hamilton 1960), Center for Epidemiological Studies Depression Scale (Sayer Radloff 1977), is eliminated. Despite the use of different items upon repeat testing within individuals, the CAT-MH also has higher test-retest reliability ($r = 0.92$) relative to traditional fixed-length tests such as the PHQ-9 ($r = 0.84$) (Beiser and Gibbons 2016).

In this study, we used CAT-MH to examine the degree to which the experience of perinatal mood disorders differs from nonpregnant adult psychiatric patients in order to determine whether a variable measures strategy is appropriate as a diagnostic screener in this population.

Materials and methods

Subjects

Psychiatric base sample

Study participants were male and female treatment-seeking outpatients between 18 and 80 years of age. Patients were recruited from two facilities, the Western Psychiatric Institute and Clinic (WPIC) at the University of Pittsburgh and a community clinic at DuBois Regional Medical Center (DuBois RMC). Psychiatric diagnoses were confirmed by medical records and the treating physician or clinician. Patients with and without a lifetime diagnosis of major depressive disorder (MDD) were included. Individuals with schizophrenia, schizoaffective disorder, or psychosis, organic neuropsychiatric syndromes (e.g., Alzheimer disease), drug or alcohol dependence within the past 3 months, inpatient treatment status, and individuals who were unable or unwilling to provide informed consent were excluded. Complete details of the sample have been previously described (Gibbons 2012). Approval for the psychiatric base sample portion of the study was obtained from the respective Institutional Review Boards (IRBs), and all subjects provided informed consent.

Perinatal sample

The enrolled subjects in the current study were women receiving outpatient obstetric care between October 2014 and December 2015 in one of three settings, a high-risk maternal-fetal medicine clinic (MFM), a resident-staffed clinic serving low-income women with both low-risk and high-risk pregnancies (Community Health Center, or CHC), and a community-based general obstetric practice. All three study sites are part of a single Midwestern US academically affiliated healthcare system where universal perinatal depression screening using the EPDS has been utilized for over a decade (Gordon et al. 2006; Kim et al. 2009). Patients <18 years of age and individuals speaking insufficient English to provide informed consent and complete study procedures were excluded. Approval for the perinatal sample was provided by

the NorthShore University HealthSystem IRB, and all subjects provided informed consent. Characteristics of the sample are summarized in Table 1.

Item bank

The item bank contained 1008 items related to depression ($n=452$), mania ($n=89$), and anxiety ($n=467$). A key step in creating the original item bank (Gibbons 2012) was qualitative review of the items done by consensus among the members of the Pittsburgh research site. The items were selected based on a review of more than 100 existing depression or depression-related rating scales. Items were modified to refer to the previous 2-week period and to have consistent response categories. The majority of items were rated on a 5-point ordinal scale. Example items are provided in the online supplement of the previously published paper (Gibbons 2012).

Design

Study subjects were administered the existing CAT-MH scales which include a diagnostic screener for MDD, dimensional severity measures of depression, anxiety and mania, and a suicide screener based on a short form of the Columbia Suicide Severity Rating Scale (Posner et al. 2011). Participants were approached to seek informed consent at obstetric care visits during which they were routinely scheduled to complete the EPDS (24–28 weeks gestational age and 6 weeks postpartum). Those who consented to participate (89.7 % of those approached) completed the CAT-MH on either a desktop computer or a tablet device prior to leaving the clinic and after completing a pen and paper version of EPDS. For CAT-MH, a text alert system was in place to immediately notify clinical staff of any instances of suicide risk based on ideation and intent, plan or recent behavior, as a supplement to item 10 on EPDS regarding thoughts of self-harm.

Statistical methods

Based on the original bifactor model calibration for the psychiatric sample (Gibbons 2012), we scored the perinatal women's response patterns separately for depression, anxiety, and mania domains. The ordinal response data were then regressed on the estimated scores for each item using a logistic regression model. A slope of 1.0 is considered to represent the lower bound on good discrimination (factor loading equivalent of 0.5). The beta coefficient for the estimated severity score based on the original psychiatric sample calibration in the logistic regression describes the strength of association between the original calibration-based severity estimate and the probability of a category increase in the response scale for the perinatal subjects. This estimate is equivalent to the slope in the multidimensional (bifactor) IRT model for the primary dimension and can also be expressed as an odds ratio (OR) of 2.72 for slope = 1.0. As such, items with ORs < 2.72 have evidence of differential item functioning (DIF) (Holland and Wainer 1993) and do not discriminate well in perinatal women. Differences in the intercepts of the logistic regression between the two populations can be produced by either differences in the underlying means between the two populations or differences in the amount of severity it takes to shift between categories between the two populations (Woods 2013). In this analysis, our focus is on the key question of differences between the two populations in terms of the items' ability to discriminate between high and low levels of the construct (e.g., depression) of interest, adjusting for differences in overall mean severity at both the item and population levels which are absorbed in the intercept of the regression.

We tested the most commonly administered items (based on CAT) for DIF. These items had a minimum of 50 subjects responding to the item. There were 66 depression, 69 anxiety, and 53 mania items used in the DIF analyses. In addition to testing for DIF, we computed the percentage of patients screening positive for MDD and the percentage screening positive for MDD and in the moderate to high depressive severity category. Similar tabulations were performed for anxiety and mania severity thresholds.

Table 1 Characteristics of perinatal women screened using the CAT-MH

Attribute	Category	Count	Percent
Race[a]	Non-Hispanic Caucasian	210	50.1
	African-American	77	18.4
	Hispanic	71	16.9
	Asian	45	10.7
	Multiracial/other	16	3.8
Average age (min, max)[a]	31.6 (18, 51)	–	–
Insurance status[b]	Private insurance	324	90.5
	Public aid	34	9.5
Marital status[b]	Partnered	257	71.8
	Not partnered/unknown	101	28.2
Parity[b]	1	170	47.5
	2+	188	52.5
Route of delivery[b]	Vaginal	196	54.7
	Cesarean	162	45.3
Plurality[b]	Singleton	316	88.3
	Multiple	42	11.7
Gestational age at delivery[b]	Term	256	71.5
	Preterm[c]	102	28.5

[a] Based on 419 unique women who completed 500 total surveys

[b] Based on 358 unique women for whom delivery information was available

[c] Preterm birth rate is higher vs. the general population due to enrichment of the sample with participants recruited in settings providing high-risk pregnancy care

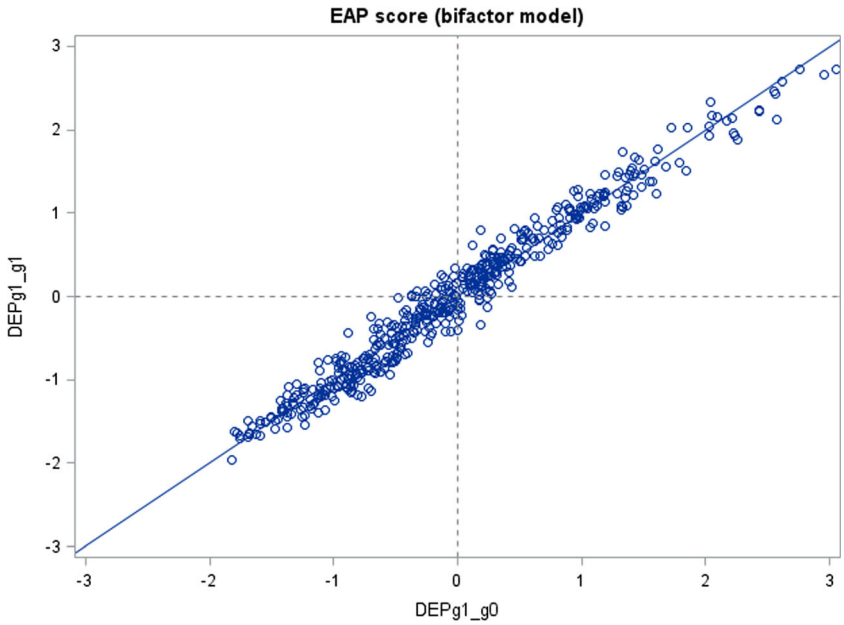

Fig. 1 Correlation between severity scores for the perinatal group (*g1*) based on the original calibration (*DEPg1_g0*) and the perinatal calibration (*DEPg1_g1*)

To examine overall test differences, we scored the perinatal data using the original bifactor model calibration based on the original psychiatric sample data and a new bifactor model calibration based on the perinatal data and computed their correlation. Differences in scale can occur if there are differences in the severity level between the two populations, which can be removed by equating the distribution of the severity scores to have mean zero and variance one. We would expect such differences because the majority of the original sample was obtained from psychiatric clinics. This latter test examines the extent to which the optimal calibration for the perinatal data produces severity estimates which differ from those based on the original calibration.

To examine convergent validity, we computed correlations of each of the CAT-MH scale scores with the EPDS. To determine concordance in terms of classification of "at-risk" status, we examined at-risk status as determined by EPDS total score and as determined by the CAD-MDD.

Results

The CAT-MH was completed by 419 women who completed a total of 500 assessments (some women participated once during pregnancy and again in the postpartum). The

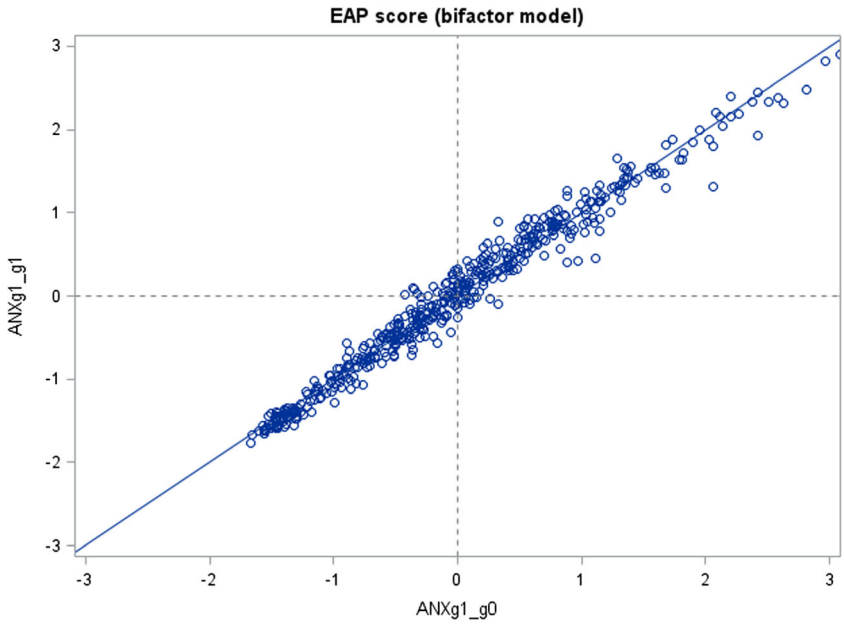

Fig. 2 Correlation between anxiety severity scores for the perinatal group (*g1*) based on the original calibration (*ANXg1_g0*) and the perinatal calibration (*ANXg1_g1*)

CAT-MH had a median assessment time of 6 min and 13 s (CAD-MDD screener 44 s, CAT-DI 75 s, CAT-ANX 76 s, CAT-MANIA 166 s, and C-SSRS 12 s). The median number of items administered was 44 items (CAD-MDD 4 items, CAT-DI 11 items, CAT-ANX 11 items, CAT-MANIA 16 items, C-SSRS 2 items).

Rates of major depressive disorder and severity

In the perinatal sample, the rate of MDD based on the CAD-MDD was 13 %, with 2 % having MDD in the moderate or severe categories (scores of 65 or greater on a 100 point scale). The rate of moderate or severe anxiety was 5 % and the rate of elevated mania was 6 %. The 0.4 % of the perinatal women screened positive for suicide risk based on ideation and intent, plan or recent behavior.

Depression, anxiety, and mania DIF

Of the 66 depression items evaluated, only 1 item exhibited DIF (i.e., failure to discriminate between high and low levels of depression) in the perinatal sample based on the psychiatric sample calibration parameters: (*In the past 2 weeks, how much have you gotten fatigued easily?*). The correlation between the depression severity scores based on the original calibration and the perinatal calibration was $r = 0.983$ (Fig. 1). Of the 69 anxiety items evaluated, only 1 item exhibited DIF (i.e., failure to discriminate between high and low levels of anxiety) in the perinatal sample based on the psychiatric sample calibration parameters: (*In the past 2 weeks, how often did you have trouble falling asleep?*) The correlation between the anxiety severity scores based on the original calibration and the perinatal calibration was $r = 0.986$ (Fig. 2). Finally, of the 53 mania items evaluated, 15 exhibited DIF (i.e., failure to discriminate between high and low levels of mania/hypomania) in the perinatal sample based on the psychiatric sample calibration parameters (Table 2). These items were generally related to symptoms that would be either commonly expected (e.g., craving sweets and carbohydrates; diminished interest in sex) or not expected (e.g., risk-taking behavior; sexual promiscuity) among perinatal women whether they are manic or not. The correlation between the mania/hypomania severity scores based on the original calibration and the perinatal calibration was only $r = 0.730$ (Fig. 3). Removal of the items that exhibited DIF increased the correlation to $r = 0.932$. The revised rates of depression, anxiety, and mania after removing the items exhibiting DIF were 4.2, 3.8, and 7.2 %, respectively.

Table 2 Items that failed to discriminate between low and high levels of mania in the perinatal population

1. In the past 2 weeks, have you had periods of at least 3 days in which you felt persistently good or high?
2. In the past 2 weeks, have you had periods of at least 3 days in which you were full of plans or got involved in many projects, jumping from one activity to another?
3. In the past 2 weeks, have you had periods of at least 3 days in which you were warm, extroverted, and sociable and it was very easy to introduce yourself to others or to make new friends?
4. In the past 2 weeks, have you had periods of at least 3 days in which you found it very pleasurable and easy to buy things, even things you didn't need?
5. In the past 2 weeks, have you had periods of at least 3 days in which you did such things as spend too much money?
6. In the past 2 weeks, have you had periods of at least 3 days in which you did such things as make foolish business decisions?
7. In the past 2 weeks, did you ever engage in risk-taking behaviors, such as driving fast, promiscuous sex, hanging out in dangerous neighborhoods?
8. In the past 2 weeks, have you been the type of person or have others told you that you usually found exciting what others would find frightening?
9. In the past 2 weeks, have you had periods of at least 3 days in which you felt as if you would like to run away from your current life, for example, by getting on the highway and driving away?
10. In the past 2 weeks, have you had periods of at least 3 days in which you were preoccupied with yourself and your own problems, thoughts, and feelings?
11. In the past 2 weeks, have you had periods of at least 3 days in which you felt that your ideas came and went unusually easily, as if your thoughts were racing?
12. In the past 2 weeks, have you had periods of at least 3 days in which you were particularly sensitive to the forms and harmony in nature?
13. In the past 2 weeks, have you had periods of at least 3 days in which you were less sexually active than is typical for you?
14. In the past 2 weeks, have you had periods of at least 3 days in which you were more interested in sex?
15. In the past 2 weeks, have you had periods of at least 3 days in which you constantly craved sweets or carbohydrates?

Convergent validity

The correlations between the modified CAT-MH and EPDS were $r = 0.82$ for depression, $r = 0.79$ for anxiety, and $r = 0.31$ for mania. Correlations between CAT-MH scale scores and EPDS total score remained similar to those for the unadjusted CAT-MH (0.82 for depression, 0.78 for anxiety, and 0.31 for mania). We also found that 91.3 % of cases were concordantly categorized as either "at-risk" or "low risk" between the two measures. Finally, concordance between suicidal ideation as assessed by EPDS item 10 and the C-SSRS items incorporated in the CAT-MH was 96.1 %.

Discussion and conclusion

Overall, there was very little evidence of DIF in depression and anxiety symptoms for the CAT-MH in perinatal women

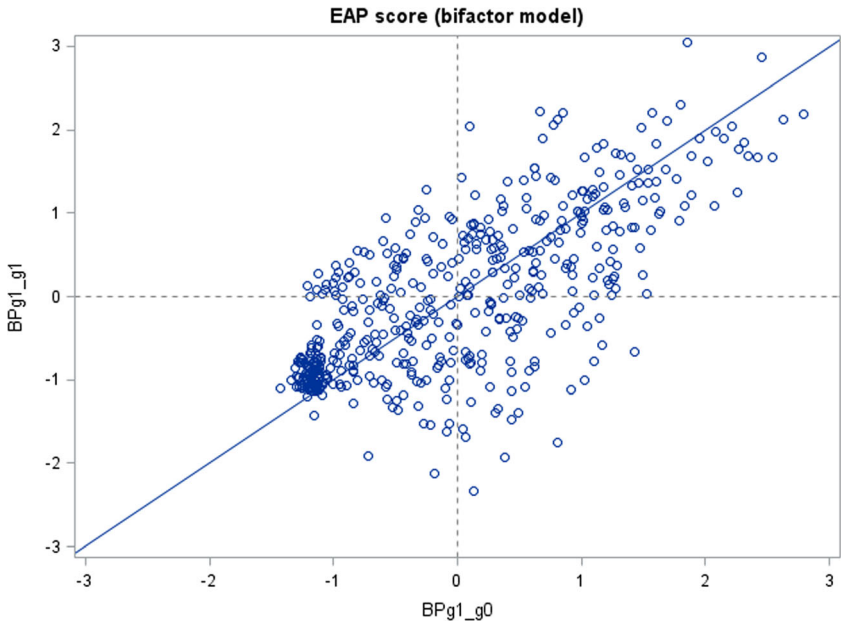

Fig. 3 Correlation between mania/hypomania severity scores for the perinatal group (*g1*) based on the original calibration (*BPg1_g0*) and the perinatal calibration (*BPg1_g1*)

compared to the original psychiatric calibration sample. This was true for the most commonly administered items for which there were sufficient data to test for DIF. Exceptions were the depression item related to fatigue and the anxiety item related to trouble falling asleep, which are both symptoms that are clearly a part of the experience of pregnancy and not unique to the experience of depression. Further, there was strong agreement between the original calibration and the perinatal-specific calibration for depression $r = 0.98$ and anxiety $r = 0.99$. The same is not true for mania, where 15 of the 53 commonly administered items (28 %) were poor discriminators in the perinatal sample. However, eliminating these items increased the correlation to $r = 0.93$ which is sufficient for routine assessment of mania/hypomania in this population. The longer assessment time and increased number of items required for the CAT-MANIA scale are likely due to the inclusion of these poor discriminating items.

While there are many different approaches to the analysis of DIF (8–10), the approach used here has several advantages for determining DIF from CAT-based testing of multidimensional constructs. First, it preserves the multidimensional nature of the underlying IRT model, whereas approaches based on multiple-group IRT (Woods et al. 2013) generally are based on unidimensional IRT and can lead to biased results. Second, the use of the logistic regression model permits DIF analyses where the number of subjects taking any particular item can be small. In our case, we had 419 subjects who completed 500 CAT-MH surveys; however, our analysis was restricted to items (symptoms) administered to only 50 or more subjects. Nevertheless, we were able to detect DIF where it existed. Third, our analysis focused on the item's ability to discriminate high and low levels of the underlying traits of interest while holding differences in population means and item parameters related to prevalence constant. The key interest here is determining which items should and should not be used in patients with a particular comorbidity, in this case, pregnancy. Overall, the major advantage is that this approach provides for continuous quality improvement where the results of routine adaptive testing in a population of interest can be used to determine DIF once a sufficient number of CAT interviews have been conducted. Here, 500 interviews produced reasonable results for DIF testing based on large item banks.

Symptom expression for many disorders can certainly differ biologically. For example, gender-specific symptoms predicting a heart attack were long ignored, leading to countless numbers of women dying from a heart attack. Symptom expression can also vary because of comorbid diagnoses or experiences which can produce or alter the expression of depressive symptomatology for reasons that are unrelated to depression. In a perinatal population, somatic symptoms that may be informative regarding depression in a psychiatric population may be less reliable differentiators when these same symptoms are either produced or moderated by pregnancy and childbirth.

The primary limitation of this study is selection of participants from a single Midwestern healthcare institution which could limit generalizability, even though our sample was ethnically diverse (50 % non-Caucasian). Likewise, 90 % of our participants were privately insured, which may affect the frequency and severity of MDD observed but should not directly influence our calibration process as compared to nonpregnant adults. As CAT-MH is deployed in both rural and urban populations with greater proportions of low-income pregnant women, it will be important to determine if there are any other DIF findings that warrant adjustment of the item bank.

Strengths of the study include our multidimensional analysis strategy for revising CAT-MH for perinatal women and our contemporaneous assessment of these patients with EPDS, the latter of which allowed for correlations to be reported. The ability to automatically track and report the time required for CAT-MH to be completed via the internet (a current feature of the software) augmented our ability to determine feasibility and to comment on the low likelihood that patients would find this to be a barrier to uptake.

Now that CAT-MH is calibrated for the perinatal population, we believe it to be an ideal diagnostic screening instrument for pregnant and postpartum women. Recognizing that pregnancy includes a risk trajectory that spans almost 2 years (i.e., conception – 1 year postpartum), CAT-MH can also be administered serially to capture newly-symptomatic patients throughout this timeframe. It may also be ideal for integration into stepped care models that rely on symptom severity to determine initial level of intervention and then monitor changing acuity levels over time to guide modification, withdrawal, or augmentation of therapy. Because CAT-MH is currently delivered via a secure mobile health application, its use can reduce barriers to screening and it can be easily integrated with electronic health records making results immediately available to clinicians even though patients can complete their evaluations remotely via cell phone and the internet. The CAT-MH can be used to screen for MDD, measure the severity of depression, and assess suicide risk in approximately 2 min. In our estimation, the extra time burden for completion of the full CAT-MH suite (6 min), while longer than fixed-item instruments like EPDS, is more than offset by both the convenience for patients in completing their screens online and on their own schedule, as well as the additional diagnostic and symptom severity data obtained.

Acknowledgments We would like to acknowledge the contributions of Kerry Swenson, MD, PhD, Lydia Muhlbach, RN, Jane Karol, Vania Blancas, Catherine Sanders, and Deanna Hanks, MHSA, FACHE, in facilitating data collection. We would like to thank Dave Patterson and Monica Jercan from Discerning Systems Inc. for programming support on customizing the CAT-MH™ for this study and providing data for the analysis. We would like to thank Yehuda Cohen from Adaptive Testing Technologies for donation of the CAT-MH™ tests. This work was previously presented in part at the 2nd Biennial Perinatal Mental Health Meeting, Chicago, IL, November 2015, and at the Biennial Meeting of the North American Society for Psychosocial Obstetrics and Gynecology, New York, NY, April 2016.

Compliance with ethical standards

Conflict of interest Robert Gibbons is a founder of Adaptive Testing Technologies that now distributes the CAT-MH™. Jo Kim, Richard Silver, Rita Elue, Marci Adams, Laura La Porte, Li Cai, and Jong Bae Kim declare that they have no conflict of interest.

Funding This study was funded by Satter Foundation.

References

American College of Obstetricians and Gynecologists (2015) Screening for perinatal depression. Committee Opinion No. 630. Obstet Gynecol 125:1268–71

Beiser D, Vu M, Gibbons R (2016) Test-retest reliability of a computerized adaptive depression test. Psychiatr Serv. doi:10.1176/appi.ps.201500304

Cogill SR, Caplan HL, Alexandra H, Robson KM, Kumar R (1986) Impact of maternal postnatal depression on cognitive development of young children. Br Med J (Clin Res Ed) 292(6529):1165–1167

Cox JL, Holden JM, Sagovsky R (1987) Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. Br J Psychiatry 150(6):782–786

Dayan J, Creveuil C, Marks MN, Conroy S, Herlicoviez M, Dreyfus M, Tordjman S (2006) Prenatal depression, prenatal anxiety, and spontaneous preterm birth: a prospective cohort study among women with early and regular care. Psychosom Med 68(6):938–946

deGruy F, Pincus H (1996) The DSM-IV-PC: a manual for diagnosing mental disorders in the primary care setting. J Am Board Fam Med 9(4):274–281

Earls MF (2010) Incorporating recognition and management of perinatal and postpartum depression into pediatric practice. Pediatrics 126(5): 1032–1039

Evins GG, Theofrastous JP, Galvin SL (2000) Postpartum depression: a comparison of screening and routine clinical evaluation. Am J Obstet Gynecol 182(5):1080–1082

Georgiopoulos AM, Bryan TL, Wollan P, Yawn BP (2001) Routine screening for postpartum depression. J Fam Pract 50(2):117–117

Gibbons RD, Hedeker DR (1992) Full-information item bi-factor analysis. Psychometrika 57(3):423–436

Gibbons RD, Weiss DJ, Pilkonis PA, Frank E, Moore T, Kim JB, Kupfer DJ (2012) Development of a computerized adaptive test for depression. Arch Gen Psychiatry 69(11):1104–1112

Gibbons RD, Hooker G, Finkelman MD, Weiss DJ, Pilkonis PA, Frank E, Kupfer DJ (2013) The computerized adaptive diagnostic test for major depressive disorder (CAD-MDD): a screening tool for depression. J Clin Psychiatry 74(7):1–478

Gibbons RD, Weiss DJ, Pilkonis PA, Frank E, Moore T, Kim JB, Kupfer DJ (2014) Development of the CAT-ANX: a computerized adaptive test for anxiety. Am J Psychiatry 171(2):187–94

Gordon TE, Cardone IA, Kim JJ, Gordon SM, Silver RK (2006) Universal perinatal depression screening in an Academic Medical Center. Obstet Gynecol 107(2, Part 1):342–347

Hamilton M (1960) A rating scale for depression. J Neurol, neurosurgery, and psychiatr 23(1):56

Hatton DC, Harrison-Hohner J, Coste S, Dorato V, Curet LB, McCarron DA (2005) Symptoms of postpartum depression and breastfeeding. J Hum Lact 21(4):444–449

Heneghan AM, Silver EJ, Bauman LJ, Stein RE (2000) Do pediatricians recognize mothers with depressive symptoms? Pediatrics 106(6): 1367–1373

Hirst KP, Moutier CY (2010) Postpartum major depression. Women 100: 17–19

Holland P, Wainer H (1993) Differential Item Functioning. Lawrence Erlbaum Associates, Hillside (NJ)

Kim JJ, La Porte LM, Adams MG, Gordon TE, Kuendig JM, Silver RK (2009) Obstetric care provider engagement in a perinatal depression screening program. Arch Womens Ment Health 12(3):167–172

Kroencke K, Spitzer R, Williams J (2001) The phq-9: validity of a brief depression severity measure. J Gen Intern Med 16(9):606–13

Kurki T, Hiilesmaa V, Raitasalo R, Mattila H, Ylikorkala O (2000) Depression and anxiety in early pregnancy and risk for preeclampsia. Obstet Gynecol 95(4):487–490

Posner K, Brown GK, Stanley B, Brent DA, Yershova KV, Oquendo MA, Mann JJ (2011) The Columbia–Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. Am J Psychiatry 168(12):1266–77

Radloff LS (1977) The CES-D scale a self-report depression scale for research in the general population. Appl Psychol Meas 1(3):385–401

Rahman A, Iqbal Z, Bunn J, Lovel H, Harrington R (2004) Impact of maternal depression on infant nutritional status and illness: a cohort study. Arch Gen Psychiatry 61(9):946–952

Siu AL, Bibbins-Domingo K, Grossman DC, Baumann LC, Davidson KW, Ebell M, Krist AH (2016) Screening for depression in adults: US Preventive Services Task Force recommendation statement. JAMA 315(4):380–387

van Doesum KT, Hosman CM, Riksen-Walraven JM, Hoefnagels C (2007) Correlates of depressed mothers' sensitivity toward their infants: the role of maternal, child, and contextual characteristics. J Am Acad Child Adolec Psychiatry 46(6):747–756

Weiss DJ (1985) Adaptive testing by computer. J Consult Clin Psychol 53(6):774

Whitaker RC, Orzol SM, Kahn RS (2006) Maternal mental health, substance use, and domestic violence in the year after delivery and subsequent behavior problems in children at age 3 years. Arch Gen Psychiatry 63(5):551–560

Woods CM, Cai L, Wang M (2013) The Langer-improved Wald test for DIF testing with multiple groups evaluation and comparison to two-group IRT. Educ Psychol Meas 73(3):532–547