Received: 22 December 2021 | Revised: 17 April 2022 | Accepted: 21 April 2022

DOI: 10.1002/brb3.2614

Brain and Behavior Open Access WILEY

COMMENTARY

Maternal major depression disorder misclassification errors: Remedies for valid individual- and population-level inference

Arthur H. Owora

Department of Epidemiology and Biostatistics, School of Public Health, Indiana University, Bloomington, Indiana

Correspondence
Arthur H. Owora, Department of Epidemiology and Biostatistics, School of Public Health, Indiana University, 1025 E 7th St #111, Bloomington, IN 47405, USA.
Email: ahowora@iu.edu

Funding information
Gallahue Family Professorship in Child Development School of Public Health-Bloomington

Abstract

Individual and population level inference about risk and burden of MDD, particularly maternal MDD, is often made using case-finding tools that are imperfect and prone to misclassification error (i.e. false positives and negatives). These errors or biases are rarely accounted for and lead to inappropriate clinical decisions, inefficient allocation of scarce resources, and poor planning of maternal MDD prevention and treatment interventions. The argument that the use of existing maternal MDD case-finding instruments results in misclassification errors is not new; in fact, it has been argued for decades, but by and large its implications and particularly how to correct for these errors for valid inference is unexplored. Correction of the estimates of maternal MDD prevalence, case-finding tool sensitivity and specificity is possible and should be done to inform valid individual and population-level inferences.

KEYWORDS
inference, maternal MDD, MDD case-finding tools, misclassification error

1 | MAJOR DEPRESSION DISORDER DEFINITION

Major depression disorder (MDD, also known as clinical depression) is a mental health condition characterized by a depressed mood and/or anhedonia (loss of interest or pleasure), and at least five of seven other symptoms that reflect a change in normal functioning or impaired functioning such as feelings of worthlessness or guilt, insomnia or hypersomnia, irritability, low energy levels, changes in appetite and weight, reduced concentration and suicidal thoughts almost all day, every day for at least a 2-week period (American Psychiatric Association, 2000). This definition of MDD is symptom-based and could be viewed as an abstract concept; making it impossible to diagnose MDD objectively. The diagnosis of MDD is further complicated by its unclear etiology and pathophysiology (Hasler, 2010).

2 | ETIOLOGY AND PATHOPHYSIOLOGY

MDD is thought to be caused by complex processes involving biological, psychological, and social factors (Dowrick, 2013; Garcia-Toro & Aguirre, 2007; Kendler et al., 1993; Schotte et al., 2006). Existing observational research shows some evidence linking genetics (Ising & Holsboer, 2006), immunological (Segerstrom & Miller, 2004), hormonal, neurological, and neuroendocrinological body mechanisms to stress response, an important etiologic risk factor of MDD (Kendler et al., 1993; Wang, 2005). Evidence regarding the functional changes in the brain causally related with depressive symptoms is mixed (Hasler, 2010). It has been suggested that multiple factors, such as genetic vulnerability (Sullivan et al., 2000), altered hypothalamic–pituitary–adrenal (HPA) axis function (Dantzer et al., 2008), deficiency of

monoamines (Kirsch et al., 2002; Speerforck et al., 2014), dysfunction of specific brain regions (Kerestes et al., 2014; Maletic et al., 2007), neurotoxic and neurotrophic processes (Serafini et al., 2014), altered glutamatergic and GABAergic neurotransmission (Hasler et al., 2007), dysregulation of glutamate system (Hasler et al., 2007; McEwen et al., 2012), and impaired circadian rhythms (Golden et al., 2005; Povitz et al., 2014), may have independent and cumulative effects that mediate or moderate each other's effects to cause MDD symptoms. Some cohort studies have suggested that age at first onset (which can occur at any time) may reflect different causal mechanisms (Burke et al., 1991; Kessler et al., 2007; Weissman et al., 1988). First-time diagnosis during childhood may be indicative of genetic predisposition (Hazell, 2002a; Rice et al., 2002) or exposure to psychosocial childhood adversity (Hazell, 2002a). During adolescence, etiology has been mainly attributed to psychosocial and economic factors (Birmaher et al., 1996; Hazell, 2002b). At this age, a disparity in MDD incidence and prevalence by sex emerges with significantly higher incidence and prevalence among girls than boys (Hankin & Abramson, 2001; Nolen-Hoeksema & Girgus, 1994). Reasons for this disparity include differences in biological body mechanisms, stress sensitivity, culture, and stress coping strategies between males and females (Nolen-Hoeksema, 1991; Nolen-Hoeksema et al., 1991; Shih et al., 2006). This gender disparity in morbidity persists into adulthood. Women as a function of changing biological and hormonal factors remain at high risk of MDD during their childbearing years (Kessler, 2003; Kessler et al., 1994) particularly during prenatal and postnatal periods.

3 | EFFECTS OF MATERNAL MDD

Among pregnant women, MDD negatively affects fetus health (Chung et al., 2001; Dieter et al., 2008; Kinsella & Monk, 2009). For example, pregnant women with MDD have been shown to have a higher fetal heart rate (FHR) than women without MDD. Following vibratory stimulation tests (i.e., clinical assessments of fetal health), fetuses of pregnant women with MDD have been shown to have delayed FHR habituation while fetuses of pregnant women without MDD had startle fetus reflex and accelerated FHR or transient tachycardia (Allister et al., 2001; Sandman et al., 2003). Among pregnant women with MDD, the higher baseline FHR and delayed habituation poststimulation are associated with HPA dysregulation (linked to higher levels of glucocorticoid transfer from mother to fetus) that negatively impact fetal development (Gilles et al., 2018; Sandman et al., 2003). Higher levels of fetal glucocorticoid exposure are associated with lower birth weight and shorter gestation at delivery (Gilles et al., 2018). Similar findings supportive of a causal hypothesis were reported in a prospective cohort study examining the association between FHR and general psychosocial stress, a risk factor for maternal MDD (DiPietro et al., 1996). Overall, MDD during pregnancy is linked to increased risk of negative obstetric and neonatal outcomes such as preeclampsia, premature delivery, and low birth weight (Buss et al., 2012; Chung et al., 2001). During the postnatal period, these effects may be compounded by poor mother–child interactions and nurturing among mothers with MDD putting children at high risk of infant morbidity and mortality (as a function of either neglect or abuse), delay in meeting appropriate development milestones, and behavioral problems (Lovejoy, 1991; Surkan et al., 2012, 2014).

4 | MATERNAL MDD DETECTION AND DIAGNOSIS

Similar to a diagnosis of MDD in the general population, maternal MDD is not an objective diagnosis because it is in part based on subjective experiences and perceptions. As a consequence of its subjective nature, a number of different maternal MDD tools have been adopted for screening, case-finding, and diagnosis as well as for monitoring treatment progress (Myers et al, 2013). The operational definitions of MDD under these tools typically involve a count and weighting of symptoms that are present over a period of 1 or 2 weeks. The number of symptoms present (including their severity ratings) is used to set a threshold above which a patient meets the MDD operational definition (Gaynes et al., 2005; Myers et al., 2013; Pignone et al., 2002). Often, the diagnostic performance of these case-finding tools is confounded by different perceptions, cultures, and assessment periods—prenatal versus postnatal (Horwitz et al., 2007; Owora et al., 2016b). Indeed, maternal MDD is often under- or overdiagnosed due to the presence of symptoms that mimic those of normal prenatal and postnatal periods (Owora et al., 2016b). Heterogeneity has also been demonstrated in existing diagnostic accuracy studies (Owora et al., 2016a, b) in part due to clinical diversity (i.e., differences between study participants) and methodological diversity (i.e., differences in the measurement, timing, and definition of MDD). These differences have important implications for the validity of case-finding tools used to classify mothers as either MDD-positive or -negative. Some studies (Levis et al., 2020) have attempted to address potential misclassification by using higher cut-off values and/or redesign of self-reported questions (e.g., Edinburgh Postnatal Depression Scale) to reduce confusion between MDD and normal prenatal and postpartum symptoms, with mixed results for the reduction of false-positives and -negatives.

5 | MATERNAL MDD: INDIVIDUAL - AND POPULATION-LEVEL INFERENCE

In psychiatry, there continues to be a paucity of research on the impact of imperfect case-finding tools on individual- and population-level inference. Yet, if unaccounted for, misclassification of psychiatric disorders, such as MDD, can lead to inappropriate clinical decisions in patient care (e.g., treating or referring a patient without MDD for further diagnostic work-up or failing to do so for patients with MDD). Such misclassification may be more prevalent among nonspecialist than specialist clinicians in primary care and/or public health prevention program settings (Horwitz et al., 2007; Myers et al, 2013).

At the population level, estimation of disease burden or risk is hampered with direct implications for allocation of public health resources

and design or targeting of prevention efforts, respectively. The argument that the use of existing case-finding tools results in misclassification bias is not new; in fact, it has been argued for decades, but their implications and particularly how to correct for these errors for valid inference is unexplored.

In this perspective piece, I revisit why MDD measurement errors or bias are important to consider from a clinical and public health view using recent epidemiologic studies. This article is not intended to be a comprehensive review; rather, I have deliberately selected articles, some of which are my own (Owora & Carabin, 2018; Owora et al., 2019, 2016a, b), to illustrate how existing estimates of maternal MDD prevalence, case-finding tool sensitivity, and specificity can be used to generate accurate risk, burden, and measures of association to inform valid individual- and population-level inference.

6 | CORRECTING FOR MISCLASSIFICATION ERRORS TO MAKE INDIVIDUAL-LEVEL INFERENCE

The concept of quantifying perceptions or impressions in clinical decision making, especially regarding diagnosis and prognosis, is not new. For instance, the likelihood of a specific diagnosis (i.e., the presence or absence of disease) is particularly appealing in the absence of confirmatory diagnostic testing. To evaluate a disease hypothesis based on non-confirmatory test results, a positive predictive value (PPV) is defined as the probability of disease (e.g., MDD) given a positive test result (e.g., a *Center of Epidemiological Studies-Depression 20-item questionnaire [CESD20]* with a moderate or severe score ≥16). Conversely, a negative predictive value (NPV) is defined as the probability of no disease given a negative test result (e.g., no or few MDD-related symptoms reported on the *CESD20*).

When combined with disease prevalence or pretest probability of disease (P_D) and known test properties, such as sensitivity (*Se*) and specificity (*Sp*) using Bayes theorem, conditional probabilities (PPV and NPV) and likelihood ratios can be used to make individual-level inference about the probability of disease given a test result, that is, P (D|T).

$$\begin{aligned}\text{PPV} &= \text{True Positives [TP]} \div (\text{TP} + \text{False Positives [FP]}) \\ &= P_D Se \div (P_D Se + [1 - P_D] \cdot [1 - Sp]), \end{aligned} \quad (1)$$

$$\begin{aligned}\text{NPV} &= \text{True Negatives [TN]} \div (\text{TN} + \text{False Negatives [FN]}) \\ &= (1 - P_D)\, Sp \div ([1 - P_D]\, Sp + [P_D] \cdot [1 - Se]). \end{aligned} \quad (2)$$

Evidently, the higher the disease prevalence (P_D), the higher we expect PPV and NPV values. Moreover, the calculation of these values is expected to vary by test score cut points used to define MDD status.

On the other hand, likelihood ratios provide an intuitive and straightforward interpretation. The likelihood ratio is a ratio of two conditional probabilities—probability of a positive (or negative) test result given that the disease is present (or absent). Therein, two variants of the likelihood ratios are needed, one for if an individual's test is positive (positive likelihood ratio: LR^+) and another if an individual's test is negative (negative likelihood ratio: LR-).

$$LR^+ = P_{TP} \div P_{FP} = Se \div (1 - Sp), \quad (3)$$

$$LR- = P_{FN} \div P_{TN} = (1 - Se) \div Sp. \quad (4)$$

Applied to MDD, the post-test probability of disease (i.e., P (D|T) can be derived from the post-test odds (i.e., product of the pre-test odds and likelihood ratio) as:

$$\text{Post-test odds} = \text{pre-test odds } x \text{ likelihood ratio} \quad (5)$$

and

$$\text{Post-test probability} = \text{Post-test odds} / (1 + \text{post-test odds}), \quad (6)$$

where

$$\begin{aligned}\text{odds} &= \text{probability of having MDD } (P_D) / 1 \\ &\quad - \text{probability of having MDD } (P_D) \end{aligned} \quad (7)$$

and

$$\text{The probability of having MDD } (P_D) = \text{odds} / (1 + \text{odds}). \quad (8)$$

Assuming, *Se* is 80% and *Sp* is 90% and a prevalence of MDD in a hypothetical population is 20%, the LR+ = 8 (i.e., 0.8/[1–0.9] from Equation 3) and LR- = 0.2 (i.e. [1–0.8]/0.9 from Equation 4) and pretest odds = 0.25 (derived from 0.2/[1–0.2] from Equation 7), two scenarios may be developed to illustrate the use of the likelihood ratio.

Scenario one: If a woman tested positive for MDD based on a case-finding tool at a defined score threshold, her post-test odds of having MDD would be $0.25 \times 8 = 2$ (*substitution into Equation* **5**: pretest odds × LR^+).

Substitution into Equation ***6***: Post-test probability of having MDD given a positive test (i.e.,$P(D^+|T^+) = 2/[1 + 2] = 0.67(67\%)$.

These results indicate that after testing positive for MDD, a woman's probability of having a MDD diagnosis is increased by 47% (i.e., from 20%—the population prevalence). While being mindful of other conditions that may mimic, cause, or coexist with MDD, a higher post-test probability would warrant further testing to confirm or rule out a MDD diagnosis. However, recommendations and referral to psychotherapy and/or prescription of antidepressants as suggested by the American Psychiatric Association and American College of Obstetricians and Gynecologists is warranted (Myers et al., 2013; O'Hara & McCabe, 2013).

In addition to a positive MDD test result, if a woman was pregnant and aged 18–25 years old, her pretest probability would be higher than 20% (population prevalence) since national prevalence estimates show that pregnant women who are 18–25 years old have a MDD pretest probability of 36% (Zhou et al., 2019). Consequently, because such a

woman's pretest probability of MDD is greater than 20%, her post-test probability of having MDD given a positive test result will be 15% higher than that for a woman from the general population (i.e., 82% derived by the direct application of Equations **5**–**8**).

Scenario two: If a woman's case-finding tool result is negative for MDD, her post-test odds of having MDD would be $0.25 \times 0.2 = 0.05$ (*substitution into Equation* ***5***: pretest odds $\times$ LR^-).

Substitution into Equation **6**: Post-test probability of having MDD (i.e., $P(D^+|T^-) = 0.05/[0.05 + 1] = 0.05(5\%)$

Considering the American Psychiatric Association and American College of Obstetricians and Gynecologists guidelines and recommendations, without any other stressors or risk factors (e.g., history of MDD), such a woman may not necessarily warrant additional follow-up. It should be noted, however, that the above example is only for illustration purposes and is not a substitute for a full clinical workup and differential diagnoses, but hopefully augments that process for better clinical judgment among mental healthcare providers.

It is important to note, however, that the use of the likelihood ratio approach for individual-level inference is not without its own limitations. For example, (1) a given LR+ (e.g., 10) value can be generated from different combinations of *Se* and *Sp* (e.g., 10 and 99% or 40 and 96%, respectively); (2) LRs are not linear (i.e., formula involves a division "÷" arithmetic operation); and (3) precision of high and low LRs is low. Despite these limitations, the translation of the likelihood ratio approach for individual-level inference using a nomogram (Fagan, 1975) can enhance its clinical utility. If the prevalence of disease and likelihood ratios are known, one can easily find the P (D|T) associated with a particular test result (±).

7 | CORRECTING FOR MISCLASSIFICATION ERRORS TO MAKE POPULATION LEVEL INFERENCE

7.1 | Prevalence estimation

In our recent article, we illustrate the prevalence estimation problem using results of the *CESD20* (Owora & Carabin, 2018). Based on recent meta-analysis results (Owora et al., 2016b), the *CESD20* is estimated to have on average, a sensitivity (*Se*) of 84% and specificity (*Sp*) of 78% for identifying patients with moderate or severe MDD symptoms based on a total score cut point of 16. If the *CESD20* were administered to 1000 women in a population with a "true" MDD prevalence of 10%, we expect results shown in Table 1.

In this case, the estimated (biased) prevalence of MDD would be 28.2% (282/1000) which is 18.2% higher than the "true" prevalence (misclassification bias or error = 28.2−10%). The PPV of 29.8% (i.e., 84 of the 282 positive tests truly have MDD) warrants a cautious interpretation of *CESD2*-positive test results to avoid the overestimation of MDD burden.

To correct for such misclassification error, if we assume T^+ represents the number of individuals who test positive for MDD using the *CESD20* and D^+ represents the number of individuals who truly have MDD then the conditional probability of an individual testing positive given that an individual truly has MDD is equal to the probability of truly having MDD and testing positive divided by the probability of truly having MDD denoted as:

$$\begin{aligned} P\left(T^+|D^+\right) &= P\left(D + nT+\right)/P\left(D^+\right)P\left(T^+|D^+\right) \\ &= P\left(D^+|T^+\right)P\left(T^+\right)/P\left(D^+\right). \end{aligned} \tag{9}$$

Using Bayes' theorem, if we assume a gold standard test for MDD exists, we can describe the association between the observed and true status as follows:

$$P\left(T^+\right) = P\left(T^+|D^+\right)P\left(D^+\right) + P\left(T^+|D^-\right)P\left(D^-\right), \tag{10}$$

where $P(T^+)$ corresponds to the proportion of individuals testing positive for MDD (observed prevalence), $P\left(T^+|D^+\right)$ corresponds to the sensitivity of the test (*Se*), $P\left(T^+|D^-\right)$ corresponds to one minus the specificity of the test (*1-Sp*), and $P\left(D^+\right)$ to the true prevalence of MDD.

$$\begin{aligned} P\left(D^+\right) = &\left(P\left(T^+|D^+\right) + P\left(T^-|D^+\right)\right) / \left(P\left(T^+|D^+\right) + P\left(T^+|D^-\right)\right. \\ &\left. + P\left(T^-|D^-\right) + P\left(T^-|D^+\right)\right). \end{aligned} \tag{11}$$

In our previous study (Owora & Carabin, 2018), we extend these concepts in a Bayesian latent class model to demonstrate that ignoring the misclassification error of case-finding tools (e.g., *CESD20*) when estimating MDD prevalence among pregnant and postpartum women can result in an underestimation of the true MDD prevalence with misclassification bias (i.e., difference between adjusted and observed prevalence estimates) ranging from 6 to 43%, depending on the distribution of pre- versus postnatal assessments. Such bias can lead to the misappropriation of scarce resources to tackle the issue of MDD among mothers.

7.2 | Risk factor measures of association

As an extension to the above discussion, unbiased measures of association between MDD and suspected risk factors are critical to identifying modifiable factors upon which preventive interventions can be developed. In our companion article (Owora et al., 2019), we demonstrate that adjustment for misclassification error in risk association studies is possible with a direct extension of the estimation concepts covered above to incorporate comparison of misclassification error-adjusted prevalence estimates between risk factor (E^+ or E^-) categories (from Equation **11** above: $P\left(D^+\right)_{E+}/P\left(D^+\right)_{E-}$) to generate a prevalence proportion ratio. Specifically, we show that failure to adjust for case-finding tools' misclassification error can lead to the underestimation of the effects of some risk factors (e.g., intimate partner violence) or the overestimation of others (e.g., period of MDD assessment: pre- vs. postnatal) on maternal MDD with varying magnitudes depending on the overall demographic and clinical profile of an investigated study sample.

TABLE 1 Contingency table for the "true" and observed results derived from the CESD20

		True MDD Diagnosis		
		Positive	Negative	
CESD20 RESULTS	Positive (≥16)	84 (TP)	198 (FP)	282 (T^+)
	Negative (<16)	16 (FN)	702 (TN)	718 (T^-)
		100 (D^+)	900 (D^-)	

$P_{TP} = P(T^+|D^+) \times P(D^+) = Se \times P(D^+)$.
$P_{FP} = P(T^+|D^-) \times P(D^-) = (1 - Sp) \times (1 - P(D^+))$.
$P_{FN} = P(T^-|D^+) \times P(D^+) = (1\text{-}Se) \times P(D^+)$.
$P_{TN} = P(T^-|D^-) \times P(D^-) = Sp \times (1 - P(D^+))$.

TABLE 2 Contingency table for the hypothetical observed CESD20 test results by pregnancy status

		Pregnant?		
		Yes	No	
Observed CESD20	Positive (≥16)	60 (a)	40 (b)	**100**
	Negative (<16)	400 (c)	500 (d)	**900**
		460 (E^+)	**540 (E^-)**	

TABLE 3 Contingency table for the adjusted CESD20 test results by pregnancy status

		Pregnant?		
		Yes	No	
Corrected CESD20	Positive (≥16)	36 (A)	3 (B)	**39**
	Negative (<16)	424 (C)	537 (D)	**961**
		460 (E^+)	**540 (E^-)**	

For illustration, if we assume a simple scenario with nondifferential misclassification of MDD among pregnant and nonpregnant women (i.e., *Se* and *Sp* are the same for both pregnant and nonpregnant women at 84 and 93%, respectively). In a hypothetical study sample (Table 2) with 1000 women, where 46% are pregnant, and 10% get a positive CESD20 test result, the observed risk of MDD would be 88% higher among the pregnant than nonpregnant women (risk ratio: 1.8; 95%CI: 1.2, 2.6).

Applying our knowledge of *Se* and *Sp*, we can calculate adjusted estimates for each cell using the Equations 12 and 13 to generate adjusted results in Table 3, with an adjusted risk ratio of 14.1 (95% CI: 4.4–45.4)

$$A = \left(a - E^+\left[1 - Sp\right]\right) / \left(Se - \left[1 - Sp\right]\right), \quad (12)$$

where $C = E^+ - A$

$$B = \left(b - E^-\left[1 - Sp\right]\right) / \left(Se - \left[1 - Sp\right]\right), \quad (13)$$

where $C = E^- - B$.

In real life, bias can involve more the just the measurement of the outcome of interest but also exposures, confounders, mediators, or moderators; these misclassification errors can be with either nondifferential or differential (i.e., *Se* and *Sp* are different for E^+ and E^-). Moreover, these variables can be either categorical or continuous. The correction of misclassification bias (or measurement error) can involve a simple bias correction (Lash et al., 2009) to more complex approaches that include probabilistic bias correction (Fox et al., 2005), Bayesian bias-correction (MacLehose et al., 2009), modified maximum likelihood (Edwards et al., 2014), and multiple imputation (Cole et al., 2006), propensity score (Lunt et al., 2012), and/or regression calibration (Rosner et al., 1989).

In summary, interest in the validity of MDD case-finding tools among mothers of young children during the pre- and postnatal periods is well justified (Owora & Carabin, 2018; Owora et al., 2019, 2016a, b). There is a growing recognition of the multiple cross-cutting negative effects of MDD on maternal–child health during the critical developmental stages of a child (Ammerman et al., 2010; Chung et al., 2004; Heckman, 2006; Lyons-Ruth et al., 1990; Sills et al., 2007; Stewart & Vigod, 2019; Thombs et al., 2014; Whitaker et al., 2006). Errors in detection (i.e., false-positive and -negatives) can result in initiation of unnecessary treatment or failure to treat maternal MDD. Valid population-level inference related to incidence or prevalence and risk factor measures of association are critical to informing appropriate allocation of scarce healthcare resources and identifying

modifiable factors for preventive intervention, respectively. Given the availability of methods that can be used to correct for MDD misclassification errors or bias derived from imperfect case-finding tools, we recommend that the correction for these errors in clinical, public health practice, and research should be the default option, and not the exception.

CONFLICT OF INTERESTS

Arthur H. Owora (author) has no conflict of interest to declare.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1002/brb3.2614

ORCID

Arthur H. Owora https://orcid.org/0000-0002-4580-7428

REFERENCES

Allister, L., Lester, B. M., Carr, S., & Liu, J. (2001). The effects of maternal depression on fetal heart rate response to vibroacoustic stimulation. *Developmental Neuropsychology*, *20*(3), 639–651. https://doi.org/10.1207/s15326942dn2003_6

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th Text Revision ed.). APA.

Ammerman, R. T., Putnam, F. W., Bosse, N. R., Teeters, A. R., & Van Ginkel, J. B. (2010). Maternal depression in home visitation: a systematic review. *Aggress Violent Behav*, *15*(3), 191–200. https://doi.org/10.1016/j.avb.2009.12.002

Birmaher, B., Ryan, N. D., Williamson, D. E., Brent, D. A., Kaufman, J., Dahl, R. E., Petrel, J., & Nelson, B. (1996). Childhood and adolescent depression: a review of the past 10 years. Part I. *Journal of the American Academy of Child and Adolescent Psychiatry*, *35*(11), 1427–1439. https://doi.org/10.1097/00004583-199611000-00011

Burke, K., Burke, J. D., Jr, R., D, S., & Regier, D. A. (1991). Comparing age at onset of major depression and other psychiatric disorders by birth cohorts in five us community populations. *Archives of General Psychiatry*, *48*(9), 789–795. https://doi.org/10.1001/archpsyc.1991.01810330013002

Buss, C., Davis, E. P., Shahbaba, B., Pruessner, J. C., Head, K., & Sandman, C. A. (2012). Maternal cortisol over the course of pregnancy and subsequent child amygdala and hippocampus volumes and affective problems. *Proceedings of the National Academy of Sciences*, *109*(20), E1312–E1319. https://doi.org/10.1073/pnas.1201295109

Chung, E. K., McCollum, K. F., Elo, I. T., Lee, H. J., & Culhane, J. F. (2004). Maternal depressive symptoms and infant health practices among low-income women. *Pediatrics*, *113*(6), e523–e529.

Chung, T. K., Lau, T. K., Yip, A. S., Chiu, H. F., & Lee, D. T. (2001). Antepartum depressive symptomatology is associated with adverse obstetric and neonatal outcomes. *Psychosomatic Medicine*, *63*(5), 830–834.

Cole, S. R., Chu, H., & Greenland, S. (2006). Multiple-imputation for measurement-error correction. *International Journal of Epidemiology*, *35*(4), 1074–1081. https://doi.org/10.1093/ije/dyl097

Dantzer, R., O'Connor, J. C., Freund, G. G., Johnson, R. W., & Kelley, K. W. (2008). From inflammation to sickness and depression: when the immune system subjugates the brain. *Nature Reviews Neuroscience*, *9*(1), 46–56. https://doi.org/10.1038/nrn2297

Dieter, J. N. I., Emory, E. K., Johnson, K. C., & Raynor, B. D. (2008). Maternal depression and anxiety effects on the human fetus: preliminary findings and clinical implications. *Infant Mental Health Journal*, *29*(5), 420–441. https://doi.org/10.1002/imhj.20192

DiPietro, J. A., Hodgson, D. M., Costigan, K. A., Hilton, S. C., & Johnson, T. R. (1996). Development of fetal movement–fetal heart rate coupling from 20 weeks through term. *Early Human Development*, *44*(2), 139–151.

Dowrick, C. (2013). Depression as a culture-bound syndrome: implications for primary care. *British Journal of General Practice*, *63*(610), 229–230. https://doi.org/10.3399/bjgp13X665189

Edwards, J. K., Cole, S. R., Chu, H., Olshan, A. F., & Richardson, D. B. (2014). Accounting for outcome misclassification in estimates of the effect of occupational asbestos exposure on lung cancer death. *American Journal of Epidemiology*, *179*(5), 641–647. https://doi.org/10.1093/aje/kwt309

Fagan, T. J. (1975). Letter: Nomogram for Bayes theorem. *New England Journal of Medicine*, *293*(5), 257. https://doi.org/10.1056/nejm197507312930513

Fox, M. P., Lash, T. L., & Greenland, S. (2005). A method to automate probabilistic sensitivity analyses of misclassified binary variables. *International Journal of Epidemiology*, *34*(6), 1370–1376. https://doi.org/10.1093/ije/dyi184

Garcia-Toro, M., & Aguirre, I. (2007). Biopsychosocial model in depression revisited. *Medical Hypotheses*, *68*(3), 683–691. doi:http://doi.org/10.1016/j.mehy.2006.02.049

Gaynes, B. N., Gavin, N., Meltzer-Brody, S., Lohr, K. N., Swinson, T., Gartlehner, G., & Miller, W. C. (2005). Perinatal depression: prevalence, screening accuracy, and screening outcomes. *Evidence Report Technology Assessment (Summary)*, *119*, 1–8.

Gilles, M., Otto, H., Wolf, I. A. C., Scharnholz, B., Peus, V., Schredl, M., Sütterlin, M. W., Witt, S. H., Rietschel, M., Laucht, M., & Deuschle, M. (2018). Maternal hypothalamus-pituitary-adrenal (HPA) system activity and stress during pregnancy: Effects on gestational age and infant's anthropometric measures at birth. *Psychoneuroendocrinology*, *94*, 152–161. https://doi.org/10.1016/j.psyneuen.2018.04.022

Golden, R. N., Gaynes, B. N., Ekstrom, R. D., Hamer, R. M., Jacobsen, F. M., Suppes, T., Wisner, K. L., & Nemeroff, C. B. (2005). The efficacy of light therapy in the treatment of mood disorders: a review and meta-analysis of the evidence. *American Journal of Psychiatry*, *162*(4), 656–662. https://doi.org/10.1176/appi.ajp.162.4.656

Hankin, B. L., & Abramson, L. Y. (2001). Development of gender differences in depression: an elaborated cognitive vulnerability-transactional stress theory. *Psychological Bulletin*, *127*(6), 773–796.

Hasler, G. (2010). Pathophysiology of depression: do we have any solid evidence of interest to clinicians? *World Psychiatry*, *9*(3), 155–161.

Hasler, G., van der Veen, J. W., Tumonis, T., Meyers, N., Shen, J., & Drevets, W. C. (2007). Reduced prefrontal glutamate/glutamine and gamma-aminobutyric acid levels in major depression determined using proton magnetic resonance spectroscopy. *Archives of General Psychiatry*, *64*(2), 193–200. https://doi.org/10.1001/archpsyc.64.2.193

Hazell, P. (2002a). Depression in children. *BMJ*, *325*(7358), 229–230.

Hazell, P. (2002b). Depression in children and adolescents. *Clinical Evidence*, *7*, 307–313.

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, *312*(5782), 1900–1902. https://doi.org/10.1126/science.1128898

Horwitz, S. M., Kelleher, K. J., Stein, R. E., Storfer-Isser, A., Youngstrom, E. A., Park, E. R., O'Connor, K. G., & Hoagwood, K. E. (2007). Barriers to the identification and management of psychosocial issues in children and maternal depression. *Pediatrics*, *119*(1), e208–e218. https://doi.org/10.1542/peds.2005-1997

Ising, M., & Holsboer, F. (2006). Genetics of stress response and stress-related disorders. *Dialogues in Clinical Neuroscience*, *8*(4), 433–444.

Kendler, K. S., Kessler, R. C., Neale, M. C., Heath, A. C., & Eaves, L. J. (1993). The prediction of major depression in women: toward an integrated etiologic model. *American Journal of Psychiatry*, *150*(8), 1139–1148.

Kerestes, R., Davey, C. G., Stephanou, K., Whittle, S., & Harrison, B. J. (2014). Functional brain imaging studies of youth depression: a systematic review. *Neuroimage Clin*, *4*, 209–231. https://doi.org/10.1016/j.nicl.2013.11.009

Kessler, R. C. (2003). Epidemiology of women and depression. *Journal of Affective Disorders*, *74*(1), 5–13.

Kessler, R. C., Amminger, G. P., Aguilar-Gaxiola, S., Alonso, J., Lee, S., & Ustun, T. B. (2007). Age of onset of mental disorders: a review of recent literature. *Curr Opin Psychiatry*, *20*(4), 359–364. https://doi.org/10.1097/YCO.0b013e32816ebc8c

Kessler, R. C., McGonagle, K. A., Zhao, S., Nelson, C. B., Hughes, M., Eshleman, S., Wittchen, H. U., & Kendler, K. S. (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: results from the national comorbidity survey. *Archives of General Psychiatry*, *51*(1), 8–19. https://doi.org/10.1001/archpsyc.1994.03950010008002

Kinsella, M. T., & Monk, C. (2009). Impact of maternal stress, depression & anxiety on fetal neurobehavioral development. *Clinical Obstetrics and Gynecology*, *52*(3), 425–440. https://doi.org/10.1097/GRF.0b013e3181b52df1

Kirsch, I., Moore, T. J., Scoboria, A., & Nicholls, S. S. (2002). The emperor's new drugs: an analysis of antidepressant medication data submitted to the U.S. Food and Drug Administration. *Prevention and Treatment*, *5*(23), https://doi.org/10.1037/1522-3736.5.1.523a

Lash, T. L., Fox, M. P., & Fink, A. K. (2009). *Applying quantitative bias analysis to epidemiologic data* (pp. 192): Springer.

Levis, B., Negeri, Z., Sun, Y., Benedetti, A., & Thombs, B. D. (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*, *371*, m4022. https://doi.org/10.1136/bmj.m4022

Lovejoy, M. C. (1991). Maternal depression: Effects on social cognition and behavior in parent-child interactions. *Journal of Abnormal Child Psychology*, *19*(6), 693–706.

Lunt, M., Glynn, R. J., Rothman, K. J., Avorn, J., & Stürmer, T. (2012). Propensity score calibration in the absence of surrogacy. *American Journal of Epidemiology*, *175*(12), 1294–1302. https://doi.org/10.1093/aje/kwr463

Lyons-Ruth, K., Connell, D. B., Grunebaum, H. U., & Botein, S. (1990). Infants at social risk: maternal depression and family support services as mediators of infant development and security of attachment. *Child Development*, *61*(1), 85–98.

MacLehose, R. F., Olshan, A. F., Herring, A. H., Honein, M. A., Shaw, G. M., & Romitti, P. A. (2009). Bayesian methods for correcting misclassification: an example from birth defects epidemiology. *Epidemiology (Cambridge, MA)*, *20*(1), 27–35. https://doi.org/10.1097/EDE.0b013e31818ab3b0

Maletic, V., Robinson, M., Oakes, T., Iyengar, S., Ball, S. G., & Russell, J. (2007). Neurobiology of depression: an integrated view of key findings. *International Journal of Clinical Practice*, *61*(12), 2030–2040. https://doi.org/10.1111/j.1742-1241.2007.01602.x

McEwen, A. M., Burgess, D. T., Hanstock, C. C., Seres, P., Khalili, P., Newman, S. C., Baker, G. B., Mitchell, N. D., Khudabux-Der, J., Allen, P. S., & LeMelledo, J. M. (2012). Increased glutamate levels in the medial prefrontal cortex in patients with postpartum depression. *Neuropsychopharmacology*, *37*(11), 2428–2435. https://doi.org/10.1038/npp.2012.101

Myers, E. R., Aubuchon-Endsley, N., Bastian, L. A., Gierisch, J. M., Kemper, A. R., Swamy, G. K., Wald, M. F., McBroom, A. J., Lallinger, K. R., Gray, R. N., Green, C., & Sanders, G. D. (2013). *Efficacy and safety of screening for postpartum depression. Comparative effectiveness review 106 (Prepared by the Duke evidence-based practice center under Contract No. 290-2007-10066-I.) AHRQ Publication No. 13-EHC064-EF*. Agency for Healthcare Research and Quality.

Nolen-Hoeksema, S. (1991). Responses to depression and their effects on the duration of depressive episodes. *Journal of Abnormal Psychology*, *100*(4), 569–582.

Nolen-Hoeksema, S., & Girgus, J. S. (1994). The emergence of gender differences in depression during adolescence. *Psychological Bulletin*, *115*(3), 424–443.

Nolen-Hoeksema, S., Girgus, J. S., & Seligman, M. E. (1991). Sex differences in depression and explanatory style in children. *J Youth Adolesc*, *20*(2), 233–245. https://doi.org/10.1007/BF01537610

O'Hara, M. W., & McCabe, J. E. (2013). Postpartum depression: current status and future directions. *Annu Rev Clin Psychol*, *9*, 379–407. https://doi.org/10.1146/annurev-clinpsy-050212-185612

Owora, A. H., & Carabin, H. (2018). Impact of misclassification error in the estimation of maternal major depression disorder prevalence in home visitation programs. *Psychiatry Research*, *261*, 80–87. https://doi.org/10.1016/j.psychres.2017.12.047

Owora, A. H., Carabin, H., Garwe, T., & Anderson, M. P. (2019). Are we validly assessing major depression disorder risk and associated factors among mothers of young children? A cross-sectional study involving home visitation programs. *Plos One*, *14*(1), e0209735. https://doi.org/10.1371/journal.pone.0209735

Owora, A. H., Carabin, H., Reese, J., & Garwe, T. (2016a). Diagnostic performance of major depression disorder case-finding instruments used among mothers of young children in the United States: a systematic review. *Journal of Affective Disorders*, *201*, 185–193. https://doi.org/10.1016/j.jad.2016.05.015

Owora, A. H., Carabin, H., Reese, J., & Garwe, T. (2016b). Summary diagnostic validity of commonly used maternal major depression disorder case finding instruments in the United States: a meta-analysis. *Journal of Affective Disorders*, *205*, 335–343. https://doi.org/10.1016/j.jad.2016.08.014

Pignone, M. P., Gaynes, B. N., Rushton, J. L., Burchell, C. M., Orleans, C. T., Mulrow, C. D., & Lohr, K. N. (2002). Screening for depression in adults: a summary of the evidence for the U.S. Preventive Services Task Force. *Annals of Internal Medicine*, *136*(10), 765–776.

Povitz, M., Bolo, C. E., Heitman, S. J., Tsai, W. H., Wang, J., & James, M. T. (2014). Effect of treatment of obstructive sleep apnea on depressive symptoms: systematic review and meta-analysis. *Plos Medicine*, *11*(11), e1001762. https://doi.org/10.1371/journal.pmed.1001762

Rice, F., Harold, G., & Thapar, A. (2002). The genetic aetiology of childhood depression: a review. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *43*(1), 65–79.

Rosner, B., Willett, W. C., & Spiegelman, D. (1989). Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Statistics in Medicine*, *8*(9), 1051–1069. discussion 1071-1053. https://doi.org/10.1002/sim.4780080905

Sandman, C. A., Glynn, L., Wadhwa, P. D., Chicz-DeMet, A., Porto, M., & Garite, T. (2003). Maternal hypothalamic-pituitary-adrenal disregulation during the third trimester influences human fetal responses. *Developmental Neuroscience*, *25*(1), 41–49. https://doi.org/10.1159/000071467

Schotte, C. K. W., Van Den Bossche, B., De Doncker, D., Claes, S., & Cosyns, P. (2006). A biopsychosocial model as a guide for psychoeducation and treatment of depression. *Depression and Anxiety*, *23*(5), 312–324. https://doi.org/10.1002/da.20177

Segerstrom, S. C., & Miller, G. E. (2004). Psychological stress and the human immune system: a meta-analytic study of 30 years of inquiry. *Psychological Bulletin*, *130*(4), 601–630. https://doi.org/10.1037/0033-2909.130.4.601

Serafini, G., Hayley, S., Pompili, M., Dwivedi, Y., Brahmachari, G., Girardi, P., & Amore, M. (2014). Hippocampal neurogenesis, neurotrophic factors and depression: Possible therapeutic targets? *CNS Neurol Disord Drug Targets*, *13*, 1708–1721.

Shih, J. H., Eberhart, N. K., Hammen, C. L., & Brennan, P. A. (2006). Differential exposure and reactivity to interpersonal stress predict sex differences in adolescent depression. *Journal of Clinical Child and Adolescent Psychology*, *35*(1), 103–115. https://doi.org/10.1207/s15374424jccp3501_9

Sills, M. R., Shetterly, S., Xu, S., Magid, D., & Kempe, A. (2007). Association between parental depression and children's health care use. *Pediatrics*, *119*(4), e829–e836. https://doi.org/10.1542/peds.2006-2399

Speerforck, S., Schomerus, G., Pruess, S., & Angermeyer, M. C. (2014). Different biogenetic causal explanations and attitudes towards persons with major depression, schizophrenia and alcohol dependence: Is the concept of a chemical imbalance beneficial? *Journal of Affective Disorders*, *168*, 224–228. https://doi.org/10.1016/j.jad.2014.06.013

Stewart, D. E., & Vigod, S. N. (2019). Postpartum depression: pathophysiology, treatment, and emerging therapeutics. *Annual Review of Medicine*, *70*(1), 183–196. https://doi.org/10.1146/annurev-med-041217-011106

Sullivan, P. F., Neale, M. C., & Kendler, K. S. (2000). Genetic epidemiology of major depression: review and meta-analysis. *American Journal of Psychiatry*, *157*(10), 1552–1562.

Surkan, P. J., Ettinger, A. K., Ahmed, S., Minkovitz, C. S., & Strobino, D. (2012). Impact of maternal depressive symptoms on growth of preschool- and school-aged children. *Pediatrics*, *130*(4), e847–e855. https://doi.org/10.1542/peds.2011-2118

Surkan, P. J., Ettinger, A. K., Hock, R. S., Ahmed, S., Strobino, D. M., & Minkovitz, C. S. (2014). Early maternal depressive symptoms and child growth trajectories: a longitudinal analysis of a nationally representative US birth cohort. *Bmc Pediatrics [Electronic Resource]*, *14*, 185. https://doi.org/10.1186/1471-2431-14-185

Thombs, B. D., Arthurs, E., Coronado-Montoya, S., Roseman, M., Delisle, V. C., Leavens, A., …, & Zelkowitz, P. (2014). Depression screening and patient outcomes in pregnancy or postpartum: a systematic review. *Journal of Psychosomatic Research*, *76*(6), 433–446. https://doi.org/10.1016/j.jpsychores.2014.01.006

Wang, J. (2005). Work stress as a risk factor for major depressive episode(s). *Psychological Medicine*, *35*(6), 865–871.

Weissman, M. M., Warner, V., Wickramaratne, P., & Prusoff, B. A. (1988). Early-onset major depression in parents and their children. *Journal of Affective Disorders*, *15*(3), 269–277.

Whitaker, R. C., Orzol, S. M., & Kahn, R. S. (2006). Maternal mental health, substance use, and domestic violence in the year after delivery and subsequent behavior problems in children at age 3 years. *Archives of General Psychiatry*, *63*(5), 551–560. https://doi.org/10.1001/archpsyc.63.5.551

Zhou, J., Ko, J. Y., Haight, S. C., & Tong, V. T. (2019). Treatment of substance use disorders among women of reproductive age by depression and anxiety disorder status, 2008–2014. *Journal of Womens Health (2002)*, *28*(8), 1068–1076. https://doi.org/10.1089/jwh.2018.7597

How to cite this article: Owora, A. H. (2022). Maternal major depression disorder misclassification errors: Remedies for valid individual- and population- level inference. *Brain and Behavior*, *12*, e2614. https://doi.org/10.1002/brb3.2614