

ORIGINAL RESEARCH

Problems with a Diagnostic Depression Interview in a Postpartum Depression Trial

Dwenda Gjerdingen, MD, MS, Patricia McGovern, PhD, MPH, and Bruce Center, PhD

Background: Diagnostic and Statistical Manual (DSM) IV–based depression interviews, valued for their diagnostic accuracy, are often considered to be essential for depression treatment trials. However, this requirement can be problematic because of participant burden. The purpose of this article is to describe our experience with the depression component of the Structured Clinical Interview for DSM Disorders (SCID) in a postpartum depression treatment trial.

Methods: In this prospective cohort study of 506 mothers of infants from 7 primary care clinics, participants were asked to complete the depression module of the SCID interview soon after enrollment. They were asked to complete the 9-item Patient Health Questionnaire (PHQ-9) depression survey at 0 to 1, 2, 4, 6, and 9 months postpartum.

Results: Forty-five women (8.9%) had a positive SCID interview and 112 (22.1%) had a positive PHQ-9 during 0 to 9 months postpartum. Problems encountered when using the SCID depression interview included (1) lower than expected SCID-based rates of depression diagnosis (8.9%); (2) SCID non-completion by 75 women (14.8%); SCID noncompleters (vs completers) were younger, poorer, less educated, and more likely to be single and black (vs white); and (3) inconsistent SCID/PHQ-9 results. Nineteen women with moderately severe to severe PHQ-9 score elevations (≥ 15) had negative SCID scores; all of these women were functionally impaired. More than 90% of women with positive PHQ-9 scores reported some degree of impairment because of their depressive symptoms.

Conclusions: The requirement of a diagnostic depression interview resulted in selection bias and missed opportunities for depression diagnosis; these are problems that detract from the interview's key strength—its diagnostic accuracy. These problems should be considered when electing to use a DSM-IV–based depression interview in research. (J Am Board Fam Med 2011;24:187–193.)

Keywords: Depression, Postpartum Health, Quantitative Research

Major depressive disorder affects up to 22% of mothers during the year after delivery, according to best estimates from a recent meta-analysis.¹ The Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) states that postpartum-onset

specifier can be applied to major depression if the onset occurs within 4 weeks of delivery.² However, some experts believe that postpartum depression (PPD) may also begin later and that it often lasts for several months.^{3,4} Therefore, for purposes of this study, we will define PPD as major depressive disorder identified over the course of our study of women ranging from 0 to 9 months postpartum. Given the relatively high prevalence and duration of this serious disorder and the fact that PPD affects not only the mother but also the infant and other family members, ongoing PPD research is sorely needed.

Fundamental to depression research is the proper identification of potential participants with major depression. DSM-IV–based depression interviews have long been considered to be the gold standard for depression diagnosis in research and

This article was externally peer reviewed.

Submitted 30 August 2010; revised 22 November 2010; accepted 24 November 2010.

From the Department of Family Medicine and Community Health (DG); the School of Public Health (PM); and the Department of Educational Psychology (BC), University of Minnesota, St. Paul.

Funding: This study was funded by the National Institute of Mental Health (R34 MH072925).

Conflict of interest: none declared.

Disclaimer: The contents of this article are the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Mental Health.

Corresponding author: Dwenda Gjerdingen, MD, MS, 580 Rice Street, St. Paul, MN 55103 (E-mail: dgjerdin@umphysicians.umn.edu).

commonly continue to be used as such.^{5–8} The Structured Clinical Interview for DSM-IV Disorders (SCID)⁹ is a widely used DSM-IV–based diagnostic interview, and telephone administration of the SCID has been found to be acceptable to patients¹⁰ and to have 97.6% agreement with in-person administration for diagnosing depression (positive agreement, 50.0%; negative agreement, 97.5%).¹¹ Though having such a gold standard is necessary when validating depression screens, this requirement may be counterproductive for some depression treatment trials.

Potential problems that may result from requiring a formal depression interview for PPD (and other depression) treatment trials include increased study costs, the need for personnel who are trained in administration of the interview, and missed cases, either because participants could not be reached for the interview or because the interview itself may not accurately elicit depressive symptoms among some patients. For example, a recent study in Melbourne, Australia, of 168 aged-care residents with normal cognitive function found that the point prevalence of major depressive disorder rose from 16% (with the SCID alone) to 22% when an informant clinical interview was included in the diagnostic procedure. Overall, 27% of depressed residents failed to disclose symptoms during the clinical interview. It was concluded that individual interviews may be insufficient to detect depression among older adults.¹²

Although mothers of infants are not “elderly,” they may have other characteristics that would make a formal depression interview an imperfect or cumbersome diagnostic tool. For example, mothers have round-the-clock care-giving responsibilities for their dependent infants, so their schedules may be erratic and make it difficult to find uninterrupted time for an interview. They may also be threatened by the idea of verbally disclosing their depressive symptoms because of fear that they will be seen as unfit mothers and their infants will be taken away. Previous research has demonstrated this fear to be a barrier to PPD diagnosis and/or treatment.^{13,14} Indeed, even apart from such fears, some individuals may find it easier to disclose personal feelings on a written survey than verbally, as suggested by a maternal depression screening study in which higher rates of positive screens were seen with a paper-based screen than with an interview-based screen (22.9% vs 5.7%, respectively).¹⁵

The present study was part of a randomized, controlled trial testing the benefit of collaborative stepped care on PPD outcomes. Participants completed interval 9-item Patient Health Questionnaire (PHQ-9) depression surveys¹⁶ and SCIDs, and a positive SCID score was required for mothers’ randomization to treatment groups. We selected the PHQ-9 as an alternative depression measure because it includes the diagnostic criteria for depression,¹⁶ and therefore might be considered a valid diagnostic tool for future PPD studies (if validity data support this use). As reported previously, the sensitivity and specificity of the PHQ-9 in this postpartum sample were 82% and 84%, respectively, using a PHQ-9 score cutoff of >10 and the SCID as the criterion standard.¹⁷

The purpose of this study was to relate the authors’ experience with the SCID in a primary care–based sample of postpartum women. Specifically, we sought to compare rates of positive SCID and PHQ-9 scores, determine the frequency of missed SCIDs and the demographic characteristics of those with missed SCIDs, and investigate inconsistencies between SCID and PHQ-9 results. Also, we evaluated positive PHQ-9 scores against the PHQ-9 follow-up question that asks how difficult the respondent’s depressive symptoms make it to function.

Methods

Participants and Procedures

This prospective cohort study, conducted within the context of a randomized, controlled trial to test the benefit of collaborative stepped care for PPD, was approved by the University of Minnesota and North Memorial Hospital institutional review boards. Most participants were recruited during their infants’ initial well-child visits at a participating clinic from October 1, 2005, through September 30, 2006; however, 20 women who were affiliated with one of the clinics were enrolled during their maternity hospital stay). Participating clinics included 4 family medicine and 3 pediatric clinics, all located in the Minneapolis/St. Paul, Minnesota, metropolitan area. Inclusion criteria included English-literate mother (this was evaluated by telephone in questionable cases, and telephone surveys were offered to English-speaking women who preferred this option), ≥12 years of age, and having a newborn infant (0 to 1 month of age) who was registered at a participating clinic.

During the infant's initial well-child visit, mothers were informed of the study and received a consent form and initial survey, which could be completed at the time of the visit or later and returned by mail. Participants were given 2-, 4-, and 6-month follow-up surveys during subsequent well-child visits (or alternatively, they completed telephone or mailed surveys), and they received the final 9-month survey by mail.

Mothers were also asked to complete the depression module of the SCID by telephone within 2 weeks of the initial survey; if a previously nondepressed woman had a positive depression screen, they were asked to complete the depression module again later. The SCID served as our reference standard for the diagnosis of major depressive disorder, and was conducted by 3 psychology doctoral students whose training consisted of observing SCID training tapes and completing 5 practice interviews under the supervision of an experienced, doctoral-level clinical psychologist. Interviewers also had ongoing weekly group supervision throughout the study to foster consistency and to address any assessment questions or uncertainties. SCID interviewers attempted to contact participants as soon as possible after they completed their initial survey (maximum, 2-week interval), and several attempts were made to call difficult-to-reach mothers. Only participants who had positive SCID scores were formally diagnosed as depressed and randomized for the treatment trial component of the study.

Measures

Survey measures included (1) demographic characteristics (initial survey), including age, level of education, marital status, number of children, race/ethnicity, total family income, and health insurance; (2) PHQ-9 depression survey, included in all interviews and positive with a score ≥ 10 ¹⁶; (3) the PHQ-9 follow-up "difficulty" question, which assesses functional impairment by asking how difficult the depressive symptoms made it for the woman to do her work, take care of things at home, or get along with other people (responses included not at all difficult, somewhat difficult, very difficult, and extremely difficult); and (4) the telephone-administered depression module of the SCID.

Statistical Analyses

Descriptive analyses assessed participants' characteristics, the number of women who had positive

PHQ-9 and SCID scores during the 9-month course of the study, the number of women who could not be reached for an SCID, inconsistent PHQ-9/SCID results, and a record of positive PHQ-9 results before SCID-based depression diagnoses. Bivariate analyses (χ^2 and t tests) were used to compare women with positive versus negative PHQ-9 scores in their responses to the "difficulty" question, as well as a comparison of the PHQ-9 scores and demographic characteristics of women who completed and did not complete the SCID.

Results

Participants' Characteristics

A total of 506 women with infants participated in the study, which represents a response rate of approximately 33% (506 participants from 1556 eligible women), with nonresponses because of refusals to participate and mothers either ignoring or not being offered an enrollment form.¹⁷

A majority of participants were white, married, and employed (Table 1). However, compared with US norms,¹⁸ our sample had a smaller proportion of whites (67.0% vs 79.6%) and a larger proportion

Table 1. Participants' Demographic Characteristics (n = 506)

Characteristic	Value
Age (mean yr [SD])	29.1 (6.2)
Married	328 (64.8)
Have only one child	207 (41.7)
Employed (working at a job or on leave)	322 (63.6)
Race/ethnicity	
White	339 (67.0)
Black or African American	89 (17.6)
Asian	34 (6.7)
Multiracial	17 (3.4)
Other	27 (5.3)
Total family annual income	
<\$20,000	133 (27.3)
\$20,000–\$79,999	177 (36.3)
\geq \$80,000	178 (36.5)
Education	
<High school diploma	81 (16.1)
\geq 4-year degree	263 (52.2)
Health Insurance	
Private insurance	257 (55.3)
Medical assistance	142 (30.5)

Values are provided as n (%) unless otherwise indicated.

Table 2. Women with Positive Patient Health Questionnaire 9 (PHQ-9) and Structured Clinical Interview for Diagnostic and Statistical Manual (DSM) Disorders (SCID) by Interval

Depression Screen or Interview	0–1 Month (n = 506)	2 Months (n = 464)	4 Months (n = 459)	6 Months (n = 455)	9 Months (n = 472)	Total* (n = 506)
PHQ-9	63 (12.5)	33 (7.1)	32 (7.0)	23 (5.0)	48 (10.2)	112 (22.1)
SCID	20† (4.0)	9 (1.9)	4 (0.9)	6 (1.3)	6 (1.3)	45 (8.9)

Values provided as n (%).

*“Total” refers to the total number of women who had a positive result sometime between 0 and 9 months postpartum.

†Most SCID interviews were conducted at 0 to 1 month postpartum. Thereafter, SCID interviews were only performed for screen-positive women who had not previously had a positive SCID.

of blacks (17.6% vs 12.9%), Asians (6.7% vs 4.6%), women with 4-year degrees (52.2% vs 24.4%), and family incomes below the poverty threshold (27.3% vs 13%). Sixty-seven percent of participants were recruited from pediatric clinics and 33% were recruited from family medicine clinics. Thirty-four participants (6.7%) dropped out before completing the final survey. More detail about demographic characteristics was provided in a previous publication.¹⁷

Women with Positive PHQ-9 and SCID Results

During the 0- to 9-month postpartum course observed here, 112 women (22.1%) had a positive PHQ-9 (score ≥ 10), and 45 (8.9%) had a positive SCID (Table 2). The SCID was conducted an average of 7 days after the initial PHQ-9.

Missed SCIDs

A total of 75 women (14.8%) could not be reached for a SCID: 68 could not be reached initially at 0 to

1 month postpartum, and an additional 7 could not be reached at the time of a positive follow-up PHQ-9 screen. Of the 68 women who could not be reached initially for an SCID, 10 (14.7%) had a positive PHQ-9 score at the time the call was attempted.

When women who initially did not complete the SCID were compared with those who did complete the interview, those who did not complete the SCID were found to be younger, less educated, more often single and black, had lower incomes, and were more likely to be receiving medical assistance (Table 3).

Inconsistent SCID/PHQ-9 Results

Nineteen women with PHQ-9 scores of ≥ 15 (consistent with moderately severe to severe depression) did not receive a SCID-based depression diagnosis either because they could not be reached for an interview (n = 5) or they had a negative SCID score (n = 14). One of these women had a PHQ-9

Table 3. Differences in Characteristics between Women Who Completed the Structured Clinical Interview for Diagnostic and Statistical Manual (DSM) Disorders (SCID) at 0 to 1 Months and Those Who Did Not*

Characteristics	SCID Noncompleters (n = 68)	SCID Completers (n = 438)	P*
Age (mean yr)	26.4	29.6	.000
>High school education	32 (47.1)	324 (74.3)	.000
Married	25 (37.3)	303 (69.5)	.000
Race			
Black	26 (38.2)	65 (14.8)	
White	35 (51.5)	320 (73.1)	
Other	7 (10.3)	53 (12.1)	.000
Insured through medical assistance	31 (50.8)	112 (28.1)	.000
Annual family income <\$20,000	33 (50.8)	100 (23.6)	.000
PHQ-9 total score (mean)	4.4	4.5	NS

Values provided as n (%) unless otherwise indicated.

*Determined by *t* tests and χ^2 tests.

†P values for age and Patient Health Questionnaire 9 (PHQ-9) were determined by *t* tests; P values for all other variables were determined by χ^2 tests.

score of 25, with suicidal ideation at the time of a negative SCID (she was called to confirm her depressive symptoms and safety, and was advised to seek immediate help). However, 11 women with positive SCIDs had negative PHQ-9 scores at the time.

Of the 25 women who had a positive SCID after the initial 0- to 1-month interval, 9 (36%) had had positive PHQ-9 results during an earlier interval, when the SCID was either negative or not completed. Two of the 9 women had prior PHQ-9 scores that were consistent with moderately severe depression (score range, 15–19), and 2 had had prior scores that were consistent with severe depression (score range, 20–27).

Functional Impairment

Women who had positive PHQ-9 scores differed significantly from women who had negative PHQ-9 scores for their responses to the question, How difficult have these problems (depressive symptoms) made it for you to do your work, take care of things at home, or get along with other people? At each of the study intervals, >90% of women with a positive PHQ-9 score compared with 35% to 46% of women with a negative PHQ-9 score indicated that their depressive symptoms made it somewhat to extremely difficult to function ($P < .001$; Table 4).

All 19 women who had PHQ-9 scores of ≥ 15 (consistent with moderately severe or severe de-

pression) and negative SCIDs reported that their depressive symptoms made it somewhat to extremely difficult to function: 10 women found it somewhat difficult, 8 found it very difficult, and 1 found it extremely difficult to function.

Discussion

Our requirement of a formal depression diagnostic interview for depression diagnosis and randomization to treatment groups resulted in problems, including lower-than-expected depression rates, missed depression interviews, selection bias, and inconsistent PHQ-9/SCID results.

The 8.9% SCID-based depression diagnosis rate seen here was much lower than expected, and, in fact, our (significantly higher) 22% rate of positive PHQ-9 scores over 9 months more closely approximated the 22% 1-year prevalence of PPD (major depression) cited in Gaynes et al's meta-analysis.¹ This notable difference in diagnostic rates raises the question, is the SCID less palatable or convenient for new mothers than the PHQ-9, or is the difference because of a gap in predictive values of the PHQ-9 versus the SCID?

In support of the theory that the SCID might be less convenient or comfortable for mothers than the PHQ-9 is the finding that 15% of participants could not be reached for the telephone-based SCID interview, even after multiple attempts. It is very possible that this group of SCID-noncom-

Table 4. Responses of Participants to the “Difficulty” Question by Survey Interval: Number of Women with Positive Versus Negative Patient Health Questionnaire 9 Scores

Postpartum Interval (Months)	How Difficult Have These Problems (Depressive Symptoms) Made It for You To Do Your Work, Take Care of Things at Home, or Get Along With Other People?				
	1 (Not At All Difficult)	2 (Somewhat Difficult)	3 (Very Difficult)	4 (Extremely Difficult)	Sum of Columns 2–4*
0–1	6 (198)	41 (165)	12 (3)	3 (1)	56 (90.3) (169 [46.0])
2	0 (203)	19 (108)	11 (2)	3 (0)	33 (100) (110 [35.1])
4	2 (161)	16 (120)	10 (0)	4 (0)	30 (93.8) (120 [42.7])
6	1 (172)	11 (118)	7 (1)	3 (1)	21 (95.5) (120 [41.1])
9	3 (169)	25 (138)	19 (2)	1 (0)	45 (93.8) (140 [45.3])

Numbers represent the number of women with positive Patient Health Questionnaire 9 (PHQ-9) scores (total score ≥ 10). Numbers within parentheses represent the number of women with negative PHQ-9 scores. PHQ-9 positive women were significantly different from PHQ-9 negative women in their responses to the “difficulty” question at each interval ($P < .001$ by χ^2 test).

*Values provided as n (%).

pleters included missed PPD cases. For example, based on our overall 8.9% SCID-positive rate, we would have expected approximately 7 of the 75 women who did not complete the SCID to be SCID-positive had they been interviewed. This estimate is also supported by the fact that 10 of the women who did not complete the SCID had a positive PHQ-9 at the time contact was attempted. It is important to note that women who did not complete the (vs those who did) were younger, poorer, and more likely to be single and black; so, it is possible that the SCID requirement produced selection bias in the diagnosis and randomization of women to treatment groups, which eventually resulted in treatment disparities.

A number of prior studies that have used diagnostic depression interviews have not specified rates of missed depression interviews.^{19–23} However, other investigators who have included this information report SCID interview noncompletion rates of 66% to 74%,^{24,25} indicating that this has also been a problem elsewhere.

Another concern was the inconsistency between participants' SCID and PHQ-9 results. For example, 19 women who had very high PHQ-9 scores (15 to 27, representing moderately severe to severe depression) either were not recognized as depressed by the SCID interview, or the SCID affirmation of depression occurred months later. Conversely, 11 women who had positive SCIDs had a negative PHQ-9. Possible reasons for our observed PHQ-9/SCID discrepancies include inaccuracy of the PHQ-9 or SCID (one would expect greater accuracy with the SCID, our gold standard); the presence of depressive symptoms caused by other mental conditions (eg, baby "blues," bipolar disorder, subsyndromal depression, or grief); disparate timing of survey and interview (in this study, a mean of 7 days, with a maximum of 2 weeks); differences in the length of time over which symptoms were assessed (2 weeks for PHQ-9 and 1 month for SCID); interviewer technique; and mothers' level of comfort with a particular diagnostic tool or method.

It is interesting that >90% of women who had a positive PHQ-9 indicated that they had some degree of functional impairment, which speaks to the face validity of the PHQ-9 in this sample. It would be helpful if future studies addressed/confirmed the validity of the PHQ-9 plus the "difficulty" question for identifying PPD among other populations. If

the "difficulty" question were found to increase the accuracy, or at least the clinical utility, of the PHQ-9 among other populations, it may be used more routinely to help sort out women with false-positive PHQ-9 results—women who may be less likely to benefit from depression treatment.

Strengths of this study include its sample size, relative ethnic diversity, primary care base, longitudinal nature, and use of a repeated measures design when assessing PPD with the PHQ-9 and SCID. The study also has weaknesses. Although its sample was drawn from 7 family medicine and pediatric clinics, it is not demographically representative of the US population, and its modest response rate (33%) may have contributed to this problem. Although our SCID interviewers were carefully trained and had ongoing weekly supervision to encourage diagnostic consistency, we did not perform formal interrater reliability testing. Additional weaknesses are the use of only a single measure of function and use of only the depression component of the SCID, which limited our diagnostic capabilities. Finally, this study does not definitively compare and validate the SCID and PHQ-9, and it is likely that the use of the PHQ-9 for diagnostic purposes would result in some false-positives or misdiagnoses that would need to be sorted out by primary care or mental health providers to avoid mistreatment. Despite these shortcomings, the study provides preliminary findings to help researchers and clinicians weigh certain risks and benefits of using a DSM-IV–based depression interview. Additional research is needed to further evaluate and compare these tools for identifying PPD.

Conclusion

Our results show that the requirement of a diagnostic interview in PPD research can be problematic because some individuals cannot be reached for an interview, resulting in missed opportunities for diagnosis, selection bias, and possible treatment disparities. In contrast, a depression survey, though perhaps less accurate, would be easier, more cost-effective, and more inclusive. Based on these results, if a positive depression diagnosis were required to initiate some form of coordinated care or increased access to other resources, exclusive use of the SCID for diagnosis would disproportionately penalize those who need this help most: the unmar-

ried, racial minorities, and the less educated and more impoverished women. These potential problems should be considered when a decision is being made about whether to use a formal DSM-IV-based interview to identify depression in research.

References

1. Gaynes BN, Gavin N, Meltzer-Brody S, et al. Perinatal depression: prevalence, screening accuracy, and screening outcomes. Evidence report/technology assessment no. 119. Rockville, MD: Agency for Healthcare Research and Quality; 2005.
2. American Psychiatric Association. Diagnostic and statistical manual of mental disorders, 4th ed. Washington, D.C.: American Psychiatric Association; 2000.
3. Gale S, Harlow BL. Postpartum mood disorders: a review of clinical and epidemiological factors. *J Psychosom Obstet Gynecol* 2003;24:257–66.
4. Wisner KL, Parry BL, Piontek CM. Postpartum depression. *N Engl J Med* 2002;347:194–9.
5. Lowe B, Kroenke K, Herzog W, Grafe K. Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). *J Affect Disord* 2004;81:61–6.
6. Zubaran C, Foresti K, Schumacher MV, et al. Validation of a screening instrument for postpartum depression in Southern Brazil. *J Psychosom Obstet Gynaecol* 2009;30:244–54.
7. Lee DT, Yip AS, Chiu HF, Leung TY, Chung TK. Screening for postnatal depression: are specific instruments mandatory? *J Affect Disord* 2001;63:233–8.
8. Bernstein IH, Wendt B, Nasar SJ, Rush AJ. Screening for major depression in private practice. *J Psychiatr Pract* 2009;15:87–94.
9. First MG, Spitzer RL, Gibbon M, Williams JB. Structured Clinical Interview for DSM-IV Axis I Disorders. Clinical version, administration booklet. New York: Biometrics Research Department, New York State Psychiatric Institute; 1997.
10. Allen K, Cull A, Sharpe M. Diagnosing major depression in medical outpatients: acceptability of telephone interviews. *J Psychosom Res* 2003;55:385–7.
11. Cacciola JS, Alterman AI, Rutherford MJ, McKay JR, May DJ. Comparability of telephone and in-person structured clinical interview for DSM-III-R (SCID) diagnoses. *Assessment* 1999;6:235–42.
12. Davison TE, McCabe MP, Mellor D. An examination of the “gold standard” diagnosis of major depression in aged-care settings. *Am J Geriatr Psychiatry* 2009;17:359–67.
13. Heneghan AM, Mercer MB, DeLeone NL. Will mothers discuss parenting stress and depressive symptoms with their child’s pediatrician? *Pediatrics* 2004;113:460–7.
14. Gjerdingen D, Crow S, McGovern P, Miner M, Center B. Stepped care treatment of postpartum depression: impact on treatment, health, and work outcomes. *J Am Board Fam Med* 2009;22:473–82.
15. Olson AL, Dietrich AJ, Prazar G, et al. Two approaches to maternal depression screening during well child visits. *J Dev Behav Pediatr* 2005;26:169–76.
16. Kroenke K, Spitzer RS, Williams JBW. Validity of a brief depression severity measure. *J Gen Intern Med* 2001;16:606–13.
17. Gjerdingen D, Crow S, McGovern P, Miner M, Center B. Postpartum depression screening at well-child visits: validity of a 2-question screen and the PHQ-9. *Ann Fam Med* 2009;7:63–70.
18. US Census Bureau. State and county QuickFacts: USA. Available from: <http://quickfacts.census.gov/qfd/states/00000.html>. Accessed October 27, 2010.
19. Blumenthal JA, Babyak MA, Doraiswamy M, et al. Exercise and pharmacotherapy in the treatment of major depressive disorder. *Psychosom Med* 2007;69:587–96.
20. Kammerer M, Marks M, Pinard C, et al. Symptoms associated with the DSM IV diagnosis of depression in pregnancy and post partum. *Arch Womens Ment Health* 2009;12:135–41.
21. Lee DT, Yip AS, Chiu HF, Chung TK. Screening for postnatal depression using the double-test strategy. *Psychosom Med* 2000;62:258–63.
22. Cooper PJ, Murray L. The impact of psychological treatments of postnatal depression on maternal mood and infant development. In: Murray L, Cooper PJ, eds. *Postpartum depression and child development*. London: Guilford Press; 1997.
23. Muzik M, Klier CM, Rosenblum KL, Holzinger A, Umek W, Katschnig H. Are commonly used self-report inventories suitable for screening postpartum depression and anxiety disorders? *Acta Psychiatr Scand* 2000;102:71–3.
24. Lee DTS, Yip ASK, Chan SSM, Tsui MHY, Wong WS, Chung TKH. Postdelivery screening for postpartum depression. *Psychosom Med* 2003;65:357–61.
25. Wittkampf K, van Ravesteijn H, Baas K, et al. The accuracy of Patient Health Questionnaire-9 in detecting depression and measuring depression severity in high-risk groups in primary care. *Gen Hosp Psychiatry* 2009;31:451–9.