Validation of the *Edinburgh Postnatal Depression Scale* (EPDS) in a sample of mothers from the 2004 Pelotas Birth Cohort Study

Validação da *Escala de Depressão Pós-natal de Edinburgo* (EPDS) em uma amostra de mães da Coorte de Nascimento de Pelotas, 2004

Iná S. Santos [1]
Alicia Matijasevich [1]
Beatriz Franck Tavares [1]
Aluísio J. D. Barros [1]
Iara Picinini Botelho [1]
Catherine Lapolli [1]
Pedro Vieira da Silva Magalhães [2]
Ana Paula Pereira Neto Barbosa [1]
Fernando C. Barros [3]

[1] Faculdade de Medicina, Universidade Federal de Pelotas, Pelotas, Brasil.
[2] Mestrado em Saúde e Comportamento, Universidade Católica de Pelotas, Pelotas, Brasil.
[3] Centro Latinoamericano de Perinatología y Desarrollo Humano, Organización Panamericana de la Salud/ Organización Mundial de la Salud, Montevideo, Uruguay.

Correspondence
A. Matijasevich
Faculdade de Medicina, Universidade Federal de Pelotas.
Av. Duque de Caxias 250, Pelotas, RS 96030-002, Brasil.
amatija@yahoo.com

Abstract

The aim of this study was to evaluate the Edinburgh Postnatal Depression Scale *(EPDS) for screening and diagnosis of postpartum depression. Three months after delivery, EPDS was administered to 378 mothers from the 2004 Pelotas Birth Cohort Study, Rio Grande do Sul State, Brazil. Up to 15 days later, mothers were re-interviewed by mental health care professionals using a semi-structured interview based on ICD-10 (gold standard). We calculated the sensitivity and specificity of each cutoff point, and values were plotted as a receiver operator characteristic curve. The best cutoff point for screening postpartum depression was ≥ 10, with 82.6% (75.3-89.9%) sensitivity and 65.4% (59.8-71.1%) specificity. For screening moderate and severe cases, the best cutoff point was ≥ 11, with 83.8% (73.4-91.3%) sensitivity and 74.7% (69.4-79.5%) specificity. For diagnosis, EPDS was valid only for prevalence of postpartum depression in the 20-25% range, with 60% PPV for the ≥ 13 cutoff point (59.5% sensitivity; 88.4% specificity). The specificities and PPVs for all cutoff points were below those reported by other authors. Small numbers and the calculation of PPV in samples with overrepresentation of cases in the majority of studies appear to account for these differences.*

Postpartum Depression; Validation Studies; Questionnaires

Background

Postpartum depression is one of the conditions that can affect childbearing women. Puerperal mothers are more vulnerable to symptoms of depression and to depressive episodes per se [1]. In the phenomenological sense, postpartum depression is similar to depression during any other period of life. However, postpartum depression can be more serious, since depression in this period can have a negative effect on the health of both the mother and the newborn [2,3,4,5], affecting the mother-child bond, infant development, and even family organization [6,7] and the child's interpersonal relations. The onset of postpartum depression happens early, between the first week and first month after delivery. Postpartum depression can compromise breastfeeding and consequently the infant's health. In extreme cases postpartum depression can even lead to infanticide. In 50% of cases postpartum depression can persist throughout the first year after delivery and become recurrent [8,9].

According to previous studies, prevalence of postpartum depression from one month to one year after delivery in the United States and Canada ranges from 8% to 26%, [8,10,11,12,13,14] and depressive symptoms can affect up to 80% of women in the postnatal period [15].

Postpartum depression is a matter of increasing concern in several countries; investments in early detection are being made for the

development of health policies for its clinical management. In 1987, Cox et al. [16] developed the *Edinburgh Postnatal Depression Scale* (EPDS) for the identification of postpartum depression, for use in clinical and research settings. EPDS is a self-administered, 10-item scale based on previously available scales (*Irritability, Depression, and Anxiety Scale* – IDA; *Hospital Anxiety and Depression Scale* – HAD; and *Anxiety and Depression Scale*) and on items devised by the authors themselves. The scale was initially compared to the Research Diagnostic Criteria (RDC). The use of EPDS is favored because of the ease and speed of its administration. This has led to its use by health care professionals in community studies, especially for the investigation of potential cases of depression. The clinical and epidemiological value of the scale have been confirmed by several validation studies carried out in different countries, with both sensitivity and specificity in the 70-85% range, depending on the cutoff point.

The present study aimed to evaluate the validity of EPDS for the diagnosis of postpartum depression three months after delivery in a sample of mothers from the 2004 Pelotas Birth Cohort Study.

Methods

A cross-sectional study was carried out during the three-month follow-up of a birth cohort in the city of Pelotas, southern Brazil, which included all births in that city in 2004 [17]. Briefly, the Pelotas 2004 birth cohort is a population-based study including all children born in the city's five hospitals. Newborns were examined and mothers interviewed during their stay in the hospital (perinatal study). At age three months, infants were visited at home for another examination. At this point mothers were re-interviewed and the EPDS questionnaire was administered.

Instrument

In order to ensure the scale's adequacy, the ten questions were initially translated into Portuguese by one of the authors (I.S.S.). Questions were then back-translated into English by an English teacher born in the United Kingdom and living in the city of Pelotas. The instrument was administered as an interview to a small number of mothers of infants up to three months of (n = 50), who did not participate in the validation study. The original version of the test and the final version of the scale in Portuguese are presented in Table 1.

In contrast to the original, self-administered format, questions were posed to mothers by a trained interviewer, as a single block and in the same order as in the original instrument, within the cohort's regular three-month follow-up interview. The decision to pose the questions to mothers verbally was based on the fact that many mothers of newborns from the cohort had little schooling and were not familiar with self-administered data collection instruments. The administration of EPDS as an interview is accepted by the instrument's authors [16] and has been used previously [18].

Sample

The present validation study was designed to detect sensitivity and specificity ≥ 80%, with a standard error of ±5, significant to the 5% level. The three-month follow-up did not include mothers whose infants died before three months. We interviewed mothers whose infants reached age three months between January 1 and March 31, 2005 (thus born from October 1 to December 31, 2004), who responded to the EPDS questionnaire at home or at the medical school, according to the cohort's three-month follow-up procedures. This sample, which included about one-fourth of all births, consisted of 886 mothers.

We used two sample selection strategies. First, all mothers scoring at least 9 points on the 30-point EPDS were included in the study. Based on the results of previous studies, we expected to find 10-15% of mothers with positive scores (about 100-150 mothers with EPDS ≥ 9). Then, a systematic 20% sample of mothers scoring < 9 was obtained by recruiting every fifth mother. All mothers selected to participate in the validation study underwent a diagnostic interview (gold standard).

For the diagnostic interview, mothers were re-interviewed at home by a mental health professional (psychiatrist, psychologist, or psychiatry resident), previously trained for the administration of the semi-structured interview. The diagnostic interview aimed to detect current or recent (previous 15 days) depressive episodes. The gold standard interview was planned to be administered 15 days after EPDS at the latest and was based on ICD-10 (*International Statistical Classification of Diseases and Related Health Problems – 10th Revision*) diagnostic criteria [19]. According to the result of this interview, mothers were classified as "normal" or "positive", the latter including those with mild, moderate, or severe episodes of depression. Mental health professionals were blinded as to mothers' EPDS scores.

Table 1

Original version and Portuguese version of the *Edinburgh Postnatal Depression Scale*.

Original version	Portuguese version
Tick the answer that best reflects how you have been feeling over the last seven days	Marque a resposta que melhor reflete como você tem se sentido nos últimos sete dias
1. **I have laughed and been able to look on the bright side of life** ( ) Yes, as usual ( ) A little less now than before ( ) Definitely less than before ( ) No, not at all	1. **Eu tenho sido capaz de rir e achar graça das coisas** ( ) Como eu sempre fiz ( ) Não tanto quanto antes ( ) Sem dúvida, menos que antes ( ) De jeito nenhum
2. **I have looked forward to the future** ( ) Yes, as usual ( ) A little less than usual ( ) A lot less than usual ( ) Not at all	2. **Eu tenho pensado no futuro com alegria** ( ) Sim, como de costume ( ) Um pouco menos que de costume ( ) Muito menos que de costume ( ) Praticamente não
3. **I have blamed myself unjustifiably when things have gone wrong** ( ) No, not at all ( ) Rarely ( ) Sometimes ( ) Yes, very often	3. **Eu tenho me culpado sem razão quando as coisas dão errado** ( ) Não, de jeito nenhum ( ) Raramente ( ) Sim, às vezes ( ) Sim, muito freqüentemente
4. **I have become anxious or worried for no good reason** ( ) Yes, very much so ( ) Yes, sometimes ( ) No, not often ( ) No, not at all	4. **Eu tenho ficado ansiosa ou preocupada sem uma boa razão** ( ) Sim, muito seguido ( ) Sim, às vezes ( ) De vez em quando ( ) Não, de jeito nenhum
5. **I have felt frightened or panicky for no good reason** ( ) Yes, very much so ( ) Yes, sometimes ( ) No, not often ( ) No, not at all	5. **Eu tenho me sentido assustada ou em pânico sem um bom motivo** ( ) Sim, muito seguido ( ) Sim, às vezes ( ) Raramente ( ) Não, de jeito nenhum
6. **I have not been able to face up to problems** ( ) Yes, I have felt incapable of facing up to problems most of the time ( ) Yes, sometimes I have not faced up to my problems as I usually would ( ) No, in the majority of cases I have been able to face up to problems relatively well ( ) No, I have been able to face up to problems as I always have	6. **Eu tenho me sentido sobrecarregada pelas tarefas e acontecimentos do meu dia-a-dia** ( ) Sim. Na maioria das vezes eu não consigo lidar bem com eles ( ) Sim. Algumas vezes não consigo lidar bem como antes ( ) Não. Na maioria das vezes consigo lidar bem com eles ( ) Não. Eu consigo lidar com eles tão bem quanto antes
7. **I have felt so bad that I have had difficulty in sleeping** ( ) Yes, most of the time ( ) Yes, sometimes ( ) No, not often ( ) No, not at all.	7. **Eu tenho me sentido tão infeliz que eu tenho tido dificuldade de dormir** ( ) Sim, na maioria das vezes ( ) Sim, algumas vezes ( ) Raramente ( ) Não, nenhuma vez
8. **I have felt sad or unwell** ( ) Yes, most of the time ( ) Yes, often ( ) Not often ( ) Not at all	8. **Eu tenho me sentido triste ou muito mal** ( ) Sim, na maioria das vezes ( ) Sim, muitas vezes ( ) Raramente ( ) Não, de jeito nenhum
9. **I have felt so sad that I have cried** ( ) Yes, most of the time ( ) Yes, often ( ) Once in a while ( ) Never	9. **Eu tenho me sentido tão triste que tenho chorado** ( ) Sim, a maior parte do tempo ( ) Sim, muitas vezes ( ) Só de vez em quando ( ) Não, nunca

(continues)

Table 1 (continued)

Original version	Portuguese version
Tick the answer that best reflects how you have been feeling over the last seven days	Marque a resposta que melhor reflete como você tem se sentido nos últimos sete dias
10. **I have thought about injuring myself**	10. **Eu tenho pensado em fazer alguma coisa contra mim mesma.**
( ) Yes, often	( ) Sim, muitas vezes
( ) Sometimes	( ) Às vezes
( ) Rarely	( ) Raramente
( ) Never	( ) Nunca

Data analysis

For each EPDS cutoff point, we calculated the sensitivity (proportion of depressed mothers according to ICD-10 criteria that were correctly identified by EPDS), specificity (proportion of non-depressed mothers correctly identified as such by EPDS), and accuracy (proportion of results correctly identified by the scale). 95% confidence intervals were determined for each of the measures. The EPDS point showing simultaneously the highest sensitivity and specificity was determined using a receiver operator characteristic (ROC) curve. Based on the sensitivity and specificity obtained for the EPDS at the cutoff points most commonly used internationally [20], the positive predictive value (proportion of true positives among all positives identified by EPDS) in simulations for populations with different postpartum depression prevalence rates was calculated.

In order to explore the performance of EPDS in a sample of high-risk mothers, a sub-sample of mothers was selected. These mothers answered positively when inquired, during the perinatal interview, about the presence of symptoms of depression, treated or untreated, or about feeling sad or depressed, always or most of the time, during the index pregnancy. The perinatal questions were formulated as follows: *During pregnancy, did you feel depressed or have any nervous condition*? (*No, Yes, treated, and Yes, untreated*) and *During the three last months of pregnancy, did you feel sad or depressed*? (*Never, sometimes, most of the time, and always*). Mothers who answered positively to both questions were considered at high risk of postpartum depression, and the validity of EPDS was tested specifically in this group.

Also investigated was the effect on EPDS performance of a change in case definition criteria, by excluding mothers with mild episodes of postpartum depression according to the gold standard. Stata 9.1 software (Stata Corp., College Station, U.S.A.) was used for all analyses.

Ethical aspects

The research protocol was approved by the Research Ethics Committee of the University of Pelotas Medical School. Since this was a nested study within the 2004 cohort and this sub-study did not involve any additional risk to the mother, the informed consent obtained was the same as requested for participation in the cohort.

Results

Only nine mothers refused to participate in the three-month follow-up, and the EPDS was administered to 886 mothers. Of these, 378 also answered the diagnostic interview (219 with score ≥ 9 and 159 with score < 9). According to the gold standard, 105 mothers showed mild, moderate, or severe episodes of depression.

Table 2 presents the characteristics of mothers included in the study. The vast majority (83.6%) had family incomes of up to three minimum wages. About 67% were aged 20-34 years, and over one-fifth (22.2%) were adolescents. Only two mothers had never attended school, whereas 15% had 1-4 years of schooling and 40% had 9 or more years. The majority of the women were white (70.9%), and 81.2% lived with a husband or partner. Slightly more than one-third (38.4%) worked outside home during pregnancy. The majority of the pregnancies were unplanned (67.2%). The prevalence of low birth weight (< 2,500 grams) and preterm births (< 37 gestational weeks) (10,8% and 16.4%, respectively), as well as the frequency of all maternal characteristics examined in the sample, with the exception of smoking during pregnancy, were statistically similar to those of the 2004 cohort as a whole (n = 4,287). The prevalence of maternal smoking was higher in the validation sample (33.6% versus 25.1%; $p < 0.001$).

Table 3 shows the sensitivity and specificity, with the respective 95% confidence intervals, for

Table 2

Characteristics of mothers included in the validation study for the Brazilian version of the *Edinburgh Postnatal Depression Scale*. Pelotas, Rio Grande do Sul State, Brazil, 2004 (n = 378).

Variables	n	%
Family income (times minimum wage)		
≤ 1	143	37.8
1.1-3	173	45.8
3.1-6	44	11.6
6.1-10	10	2.7
> 10	8	2.1
Mother's age (years)		
< 20	84	22.2
20-34	253	66.9
≥ 35	41	10.9
Schooling (years)		
0	2	0.5
1-4	56	15.0
5-8	168	44.9
≥ 9	148	39.6
White skin color	268	70.9
Living with partner	307	81.2
Primiparae	156	41.3
Smoking during pregnancy	127	33.6
Work during pregnancy	145	38.4
Planned pregnancy	124	32.8
Preterm delivery	62	16.4
Low birth weight infant	41	10.8

Table 3

Sensitivity, specificity, and 95% confidence intervals for different cutoff points in a Brazilian version of the *Edinburgh Postnatal Depression Scale*. Pelotas, Rio Grande do Sul State, Brazil, 2004 (n = 378).

Cutoff points	Sensitivity (95%CI)	Specificity (95%CI)
≥ 3	99.0 (94.8-100.0)	16.4 (12.2-21.4)
≥ 4	98.1 (93.3-99.8)	23.7 (18.8-29.2)
≥ 5	97.1 (91.9-99.4)	33.6 (28.0-39.5)
≥ 6	96.2 (90.5-99.0)	39.8 (33.9-45.8)
≥ 7	96.2 (90.5-99.0)	45.6 (39.6-51.7)
≥ 8	93.3 (86.7-97.3)	50.4 (44.3-56.4)
≥ 9	91.3 (84.2-96.0)	54.7 (48.6-60.7)
≥ 10	82.7 (74.0-89.4)	65.3 (59.4-71.0)
≥ 11	74.0 (64.5-82.1)	77.4 (72.0-82.2)
≥ 12	65.4 (55.4-74.4)	82.1 (77.1-86.5)
≥ 13	59.6 (49.5-69.1)	88.3 (83.9-91.9)
≥ 14	50.0 (40.0-60.0)	92.3 (88.5-95.2)
≥ 15	40.4 (30.9-50.5)	94.2 (90.7-96.6)
≥ 16	36.5 (27.3-46.6)	96.4 (93.4-98.2)

each of the EPDS cutoff points. As expected, sensitivity decreased progressively as the cutoff point increased, with a more pronounced decrease between the ≥ 9 and ≥ 10 cutoff points (from 91.3% to 82.6%). In contrast, specificity between these two cutoff points increased from 54.9% to 65.4%. According to the ROC curve (Figure 1), the ≥ 10 cutoff point was best for this population. The 95% confidence intervals for this cutoff point were 75.3% to 89.9% for sensitivity and 59.8% to 71.1% for specificity.

We analyzed the effect of changes in maternal postpartum depression risk profile on EPDS performance. During the perinatal interview, a total of 247 mothers reported depression, treated or untreated, or feeling sad or depressed, always or most of the time, during the index pregnancy. These women were considered as a higher risk group for postpartum depression. Of these, the gold standard identified 89 mothers (36%) with diagnosis of postpartum depression. As in the sample from the general population of mothers, the balance between sensitivity and specificity confirmed the adequacy of the ≥ 10 cutoff point, with 79.8% sensitivity (69.9-87.6%) and 53.2% specificity (45.1-61.1%), which ratified these levels as stable characteristics of the test, regardless of the disease prevalence.

The effect of changes in the prevalence of postpartum depression in the study population was observed in the predictive value of EPDS. Table 4 shows the positive predictive values for EPDS cutoff points between 10 and 14 in simulations for populations with different postpartum depression prevalence rates. Thus, for instance, if EPDS was administered as a diagnostic test with a cutoff point of ≥ 11 in a population with a postpartum depression prevalence of about 20%, the positive predictive value would be 45%. In this case, the majority of women identified by EPDS as suffering from postpartum depression would actually be false-positives. Likewise, in populations with a postpartum depression prevalence of 15%, the use of EPDS at this same cutoff point would yield a positive predictive value of only 36.6%. As expected, lower cutoff points, such as ≥ 10, when used in a population with 15% prevalence of postpartum depression, would lead to 42% of the tested population being diagnosed as suffering from postpartum depression and to a positive predictive value of 29.6%, even lower than the previous one.

The effect of changes in postpartum depression definition criteria was tested by considering as positive only mothers classified by the gold standard as showing moderate or severe episodes of depression (75 out of 378 mothers in the general population). In this scenario, the ROC curve identified ≥ 11 as the best cutoff point, with 83.8% (73.4-91,3%) sensitivity, 74.7% (69.4-79.5%) specificity, and 76.5% accuracy. Among high-risk mothers (n = 247), the gold standard identified 63 as showing moderate or severe postpartum depression. As expected, analyses within this group confirmed the ≥ 11 cutoff point, with 81% (69.1-89.8%) sensitivity, 66.3% (59-73.1%) specificity, and 70% accuracy.

Discussion

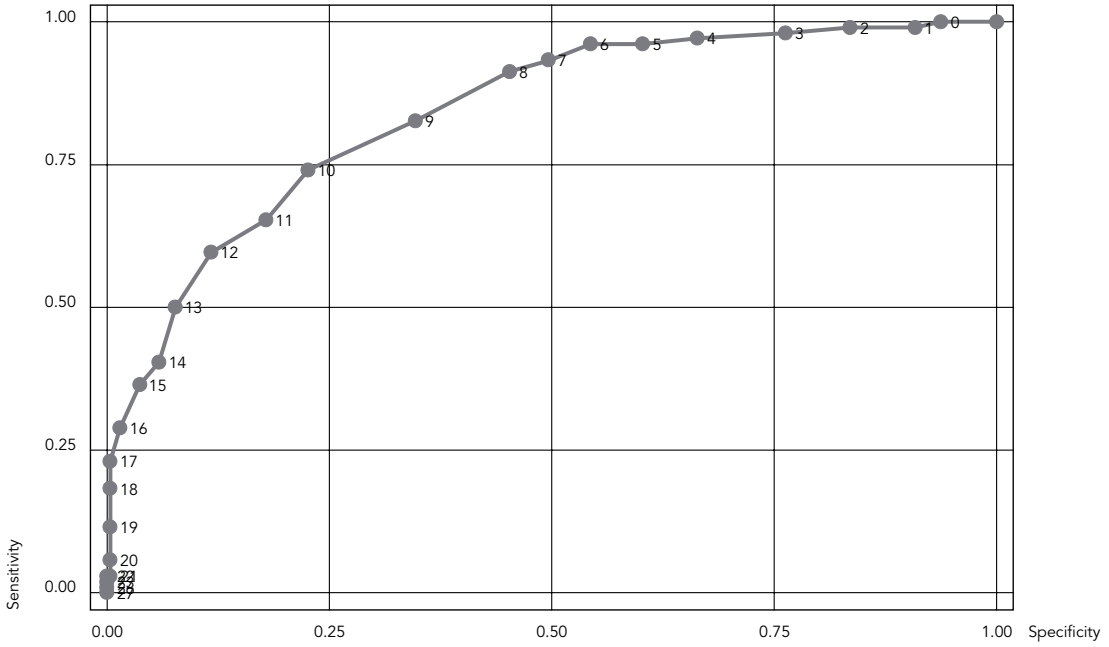
EPDS is the scale most widely used worldwide for the study of postpartum depression. It has been translated into several languages and validated in different countries. Before the present investigation, two other studies evaluated the performance of EPDS in Brazil, one in Pernambuco, in the Northeast [18] and one in the Federal District, in the Central West of the country [21]. The Pernambuco study included 218 women and aimed to measure the prevalence of pre- and postpartum depression in a sample of low-income mothers. EPDS sensitivity and specificity were evaluated only for the antenatal period, using as a gold standard the interviewers' impressions (medical and nursing students) based on the IDC-10 diagnostic criteria. Using the ≥ 13 cutoff point, EPDS showed 73% sensitivity and 90.5% specificity for diagnosing depression during the third month of pregnancy.

The validation study conducted in the Federal District included 69 predominantly middle-class working women with a mean of 10.2 weeks postpartum. According to the authors, the best cutoff point for the scale was ≥ 11, with 84% sensitivity and 82% specificity. The authors provided no information on the risk profile of the sample, but the working definition of postpartum depression included only moderate or severe episodes. This cutoff point coincides with that found by the present study when cases were defined according to the same criteria.

EPDS was originally constructed as a screening instrument for postpartum depression, but the scale's authors and others propose that, using ≥ 13 as the cutoff point, the scale has high positive predictive value for diagnosing postpartum depression. In general, EPDS validation studies report high sensitivity and specificity, as well as high positive predictive value, both as a screening instrument and as a diagnostic test. In the present study, the sensitivity of EPDS was consistent with the findings of other authors using the same cutoff points. On the other hand, specificity and positive predictive value were generally below those reported in the literature at all cutoff

Figure 1

Receiver operator characteristic curve for the performance of the *Edinburgh Postnatal Depression Scale* compared to an interview with a mental health professional using ICD-10 criteria (gold standard) for the diagnosis of postpartum depression. Pelotas, Rio Grande do Sul State, Brazil, 2004.



Area under ROC curve = 0.8401.

points investigated. The high rate of false-positives and the corresponding low specificity were largely responsible for the differences found in terms of positive predictive values.

The comparison between the results of different validation studies for a same test requires caution. In addition to the quality of the instrument used, several methodological aspects may interfere with the results obtained. These include the prevalence of the disease in the sample, the case definition employed by the gold standard, the design of the validation study, and the study population's socio-cultural characteristics. As initially described by its authors [16], EPDS was developed to screen for postpartum depression among mothers considered potentially depressed according to attending health professionals. Therefore, women with depression were over-represented in the sample. This was also the case in the present study, in which we included all mothers with EPDS ≥ 9 and 20% of those with EPDS < 9. This type of sampling design leads to a higher prevalence of postpartum depression than that observed among the general population of mothers. This methodological aspect has an effect on the test's predictive value, while sensitivity and specificity remain unchanged. A test's positive predictive value increases in populations where prevalence of the disease is greater, while sensitivity and specificity remain relatively constant. Validation studies that estimated predictive value based on samples in which mothers with postpartum depression were over-represented obtained better results than studies that corrected values according to the actual prevalence of postpartum depression. Indeed, a review of EPDS validation studies by Eberhard-Gran [20] identified positive predictive values from 37% to 78%, which, when corrected using a more realistic prevalence of 13%, were usually smaller, ranging from 22% to 79%. After correction, the positive predictive values of these studies were closer to those found in the present study for corresponding cutoff points and prevalence rates.

There are two ways of increasing a test's positive predictive value: increasing the prevalence of disease in the screened population and altering the cutoff point so as to increase the specificity. Indeed, the present study's findings indicate a higher positive predictive value when EPDS

Table 4

Positive predictive values for different *Edinburgh Postnatal Depression Scale* (EPDS) cutoff points, according to the prevalence of postpartum depression in the study population. Pelotas, Rio Grande do Sul State, Brazil, 2004.

Cutoff point	Positive predictive value	95%CI
EPDS ≥ 10		
Prevalence of postpartum depression		
5%	11.2	9.5-13.1
10%	20.9	18.1-24.2
15%	29.6	25.9-33.6
20%	37.4	33.1-41.8
25%	44.3	39.8-48.9
EPDS ≥ 11		
Prevalence of postpartum depression		
5%	14.7	11.9-18.1
10%	26.7	22.1-31.8
15%	36.6	31.1-42.5
20%	45.0	39.0-51.1
25%	52.2	46.0-58.3
EPDS ≥ 12		
Prevalence of postpartum depression		
5%	16.1	12.6-20.5
10%	28.9	23.3-35.2
15%	39.2	32.6-46.3
20%	47.8	40.6-55.0
25%	54.9	47.7-62.0
EPDS ≥ 13		
Prevalence of postpartum depression		
5%	21.2	15.8-27.8
10%	36.2	28.3-44.9
15%	47.4	38.5-56.4
20%	56.1	47.0-64.7
25%	63.0	54.2-71.0
EPDS ≥ 14		
Prevalence of postpartum depression		
5%	25.6	17.9-35.1
10%	42.0	31.5-53.3
15%	53.5	42.2-64.4
20%	62.0	50.9-72.0
25%	68.5	58.0-77.4

is applied to high-risk mothers. When the ≥ 10 cutoff was used to screen postpartum depression among a group of mothers with 25% pre-test prevalence of the disease, the positive predictive value increased from 29.6% (in the general population of mothers, with an approximate prevalence of 15%) to 44.3%. Although only two of every five mothers identified by the test as at risk of postpartum depression will actually be diagnosed with the disease when interviewed by health professionals, this is an acceptable level for screening instruments. Due to the generally low prevalence of diseases, screening tests usually have low positive predictive values, even when specificity is high [22]. On the other hand, at the ≥ 10 cutoff point, the positive predictive value of the test was too low for the scale to be recommended as a diagnostic test, even in a population with a high prevalence of postpartum depression. Despite the almost twofold increase between the pre-test (25%) and post-test (44.3%) probability of postpartum depression, the discriminating capacity of EPDS was still weak. A positive predictive value below 50% is weaker than that obtained

when tossing a coin. Furthermore, the area under the ROC curve for high-risk mothers was lower than that obtained for the entire group of mothers (0.76 and 0.84, respectively), indicating lower accuracy among high-risk mothers than among the entire group.

The second alternative to improve the predictive value of EPDS would be to choose a cutoff point with higher specificity. For instance, by increasing the cutoff point from ≥ 10 to ≥ 11, the specificity increased from 65.3% to 77.3%. This increase, however, was accompanied by a reduction in sensitivity from 82.6% to 74%. A screening test that fails to identify more than one-fourth of mothers with postpartum depression is unacceptable. A good screening test must have high sensitivity in order not to miss the few cases of the disease, and high specificity, in order to reduce the number of false-positives that will have to undergo further evaluation.

For diagnostic purposes, the best performances were found for cutoff points ≥ 13 and ≥ 14 when the test was applied to high-risk mothers, with postpartum depression prevalence between 20% and 25%. For these mothers, the positive predictive value was more than 60%.

The second methodological aspect that may interfere with the results of validation studies is the case definition used by the gold standard. In the current study, we defined postpartum depression as occurring in mothers presenting depressive episodes with any level of severity. In general, studies that include mild cases of the disease report lower sensitivity for a same cutoff point [23,24]. The current study's findings also provide evidence of this phenomenon. The sensitivity of the ≥ 10 cutoff point in the sample including all mothers with depressive episodes was 82.6%, whereas among mothers with moderate and severe episodes only, sensitivity for the ≥ 10 cutoff point increased to 87.8% (78.2-94.3%).

An alternative would be to use the ≥ 11 cutoff point to diagnose moderate or severe postpartum depression, a scenario yielding greater specificity for EPDS when compared to the diagnosis of any type of postpartum depression. Assuming 10% prevalence of moderate or severe postpartum depression, the positive predictive value for EPDS ≥ 11 would be 26.7%. For 15% prevalence, the positive predictive value would increase to 36.9%. Although a positive result in these situations would imply a probability of disease about three and two times greater, respectively, than the pre-test probability, such a finding indicates that, in the first case, only slightly more than one-fourth of mothers selected by EPDS would have moderate or severe postpartum depression confirmed after evaluation by health professionals. In the second scenario, even with a 50% increase in prevalence as compared to the previous example, only a little more than one-third of mothers would be correctly identified by EPDS. These findings indicate that a cutoff point of ≥ 11 would be adequate for screening, but not for diagnosing mothers with moderate or intense postpartum depression. For diagnostic purposes, the most adequate cutoff point would be ≥ 15, which has 47.3% (35.6-59.3%) sensitivity and 92.4% (88.9-95.1%) specificity among mothers with 20% prevalence of moderate or severe postpartum depression. A major difficulty would be to work with a sample of mothers with such high pre-test prevalence of moderate and severe postpartum depression. The accuracy (area under the ROC curve) for the diagnosis of moderate and severe cases was 0.86, thus higher than the area for the screening and diagnosis of PPD among both the general population and high-risk mothers.

Finally, study design plays an important role in the evaluation of a test's attributes. As a screening test, the aim of EPDS is to detect postpartum depression during its pre-symptomatic phase or as close as possible to the threshold of clinically detectable symptoms. When conducting a cross-sectional comparison of performance by EPDS and the psychiatric interview, the scale is actually being tested as a diagnostic test for postpartum depression rather than as a screening instrument. The sensitivity of a screening test is given by the ratio between the number of true-positives and the sum of true-positives and subjects that will develop the disease within a given follow-up period [22], meaning that the disease was present but the test was not able to identify it. Ideally, therefore, a study of the validity of EPDS as a screening test should evaluate the scale's performance in the early identification of symptoms that would later evolve to postpartum depression. A study with this aim should have a prospective design and include (as a selection criterion) only mothers tested after the first 7-10 days post-delivery, the period in which 30-70% of mothers show symptoms of melancholy, sadness, and emotional instability, which are self-limited in the majority of cases [25]. Two measurements at different time points would be necessary: one during the selection of mothers, in which EPDS would be administered, and another between four weeks and three months post-delivery, the peak of postpartum depression incidence [26], when only the gold standard for the diagnosis of postpartum depression would be administered. Mothers with positive EPDS scores at the beginning of postpartum and that developed postpartum depression during the follow-up period would be considered true-positives. Only

thus could the sensitivity of EPDS as a screening instrument be defined. Specificity, on the other hand, would be determined as the proportion of mothers with negative EPDS scores confirmed by the gold standard. Therefore, the low predictive value for EPDS found in the present study may be due to the design used for its validation. We found no studies in the literature that evaluated EPDS as a screening instrument using this methodology. The available results thus express the performance of EPDS more as a diagnostic test than as a screening instrument for postpartum depression.

The above-mentioned aspects may explain the differences detected between this and other studies with respect to the positive predictive value of EPDS. The reason for the differences in specificity and false-negative rates, on the other hand, appear to be due to other methodological aspects, especially sample size. Wide confidence intervals such as those reported by the majority of EPDS validation studies in terms of both sensitivity and specificity are due to the small samples investigated. Low specificity and the corresponding high false-positive rates found in the present study indicate that mothers answered positively to EPDS without these answers having the depressive connotation that the scale aims to detect. This characteristic is not concentrated only in a few questions; rather, it is seen in most of the test's questions. Although the translation of EPDS into Portuguese and the subsequent back-translation into English were considered adequate given the settings in which the scale would be used, it is possible that cultural factors may have interfered with the interpretation of the content of certain items and consequently with the expected values for the answers provided. Thus, before using this scale for the general population of Brazilian mothers, it would be recommendable to test the validity of the content of these items in relation to the other variables in the scale, using further studies.

Other characteristics of the study population, such as the health status of children and the mother's perception of her child's health, were not explored in the present study and may have limited the results' validity. Although the 2004 birth cohort recorded the number of hospital admissions and medical appointments during the first three months of the infant's life, the mother's perception regarding the child's health was not evaluated. However, events leading to hospitalization were infrequent in this sample. Only 20 children were hospitalized at least once before the interview date. According to the gold standard, ten of these mothers showed depressive episodes, versus 95 among the others ($p = 0.02$). Likewise, the mother's actual or self-perceived health status was not evaluated. Mothers with unfavorable self-perceived health status may show greater prevalence of postpartum depression than those who considered themselves healthy. It is plausible that mothers with clinical intercurrences due to (or increased by) pregnancy show greater prevalence of postpartum depression.

Strengths of the current study include the fact that both EPDS and the gold standard interview were standardized and blinded as to each other's results. Administration of the scale as an interview by a trained interviewer, rather than as a self-administered instrument, as originally planned and as done in the majority of studies, was appropriate for the social and educational characteristics of a population-based sample of mothers. Moreover, this was the first validation study for EPDS in Brazil to rely on a population-based sample.

Conclusions

In short, the present study has shown that the validity of EPDS should be interpreted in light of the use for which it is intended. EPDS is adequate as a screening instrument using the ≥ 10 cutoff point, especially among selected populations of mothers at high risk of postpartum depression. For diagnosis, the ≥ 13 cutoff point will be adequate only if used among high-risk populations. In the general population of mothers, the scale shows low validity for the diagnosis of postpartum depression. It should be noted, however, that there is still a gap to be filled in the validation of EPDS as a screening instrument for postpartum depression, given that with the design used by the present study and other studies identified in the literature, such performance remains to be formally tested.

Resumo

Avaliar a validade da Escala de Depressão Pós-natal de Edimburgo *(EPDS) para rastreamento e diagnóstico de depressão pós-parto. Três meses pós-parto, a EPDS foi aplicada a 378 mães da Coorte de Nascimentos de Pelotas, Rio Grande do Sul, Brasil, em 2004. Até 15 dias após, as mães foram reentrevistadas por profissionais de saúde mental utilizando-se questionário semi-estruturado baseado na CID-10 (padrão-ouro). Calculamos sensibilidade e especificidade de cada ponto de corte e construiu-se curva ROC. Melhor ponto de corte para rastreamento foi ≥ 10 (sensibilidade 82,6%, 75,3%-89,9%; especificidade 65,4%, 59,8%-71,1%). Para rastrear casos moderados e graves, melhor ponto de corte foi ≥ 11, com sensibilidade 83,8% (73,4%-91,3%) e especificidade 74,7% (69,4%-79,5%). Para diagnóstico, a EDPS foi válida somente para prevalências em torno de 20%-25%, com valor preditivo positivo de 60% para o ponto de corte ≥ 13 (sensibilidade 59,5%; especificidade 88,4%). As especificidades e valores preditivos positivos de todos os pontos de corte foram inferiores aos relatados na literatura. Possivelmente, o uso de amostras pequenas e o cálculo de valores preditivos positivos em amostras com super-representação de casos, sejam responsáveis por essas diferenças.*

Depressão Pós-Parto; Estudos de Validação; Questionários

Contributors

I. S. Santos and A. Matijasevich designed the study, conducted the data analysis, and wrote the draft and final version of the article. B. F. Tavares coordinated the fieldwork. I. P. Botelho, C. Lapolli, P. V. S. Magalhães, and A. P. P. N. Barbosa participated in the fieldwork. A. J. D. Barros and F. C. Barros helped develop various concepts and interpret the findings. All authors reviewed the draft and contributed to the final version of the article.

Acknowledgements

The authors wish to thank the funding agencies, the mothers of all newborns in the cohort, the hospitals of the city of Pelotas, the Municipal Secretariat of Health and Welfare, and all those who collaborated in the various stages of this study. The study was funded by the World Health Organization (HQ/04/072979), the Brazilian National Research Council (CNPq grant nº. 476727/2003-0), and the Children's Mission (*Pastoral da Criança*).

References

1. Buist AE, Barnett BE, Milgrom J, Pope S, Condon JT, Ellwood DA, et al. To screen or not to screen – that is the question in perinatal depression. Med J Aust 2002; 177 Suppl:S101-5.
2. Beck CT. Teetering on the edge: a substantive theory of postpartum depression. Nurs Res 1993; 42: 42-8.
3. Beck CT. The effects of postpartum depression on maternal-infant interaction: a meta-analysis. Nurs Res 1995; 44:298-304.
4. Cogil SR, Caplan HL, Alexandra H, Robson KM, Kumar R. Impact of maternal postnatal depression on cognitive development of young children. BMJ 1986; 292:1165-7.
5. Holden JM. Postnatal depression: its nature, effects, and identification using the Edinburgh Postnatal Depression Scale. Birth 1991; 18:211-21.
6. Murray J. The impact of postnatal depression on infant development. J Child Psychol Psychiatr 1992; 33:543-61.
7. Stein A, Gath D, Bucher J, Bond A, Day A. The relationship between post-natal depression and mother-child interactions. Br J Psychiatry 1992; 158:46-52.
8. Cox JL, Murray D, Chapman G. A controlled study of the onset, duration, and prevalence of postnatal depression. Br J Psychiatry 1993; 163:27-31.
9. Garcia-Esteve L, Ascaso C, Ojuel J, Navarro P. Validation of the Edinburgh Postnatal Depression Scale (EPDS) in Spanish mothers. J Affect Disord 2003; 75:71-6.
10. Sword W, Watt S, Krueger PD, Sheehan DD, Lee KS, Roberts J, et al. New mothers identify information needs: actions for postpartum community care providers. Public Health and Epidemiology Report Ontario 2000; 11:38-44.
11. Gotlib IH, Whiffen VE, Mount JH, Milne K, Cordy NI. Prevalence rates and demographic characteristics associated with depression in pregnancy and the postpartum. J Consult Clin Psychol 1989; 57:269-74.

12. O'Hara MW, Zekoski EM, Philips LH, Wright EJ. Controlled prospective study of mood disorders: comparison of childbearing and nonchildbearing women. J Abnorm Psychol 1990; 99:3-15.
13. Warner R, Appleby L, Whitton A, Faragher B. Demographic and obstetric risk factors for postnatal psychiatric morbidity. Br J Psychiatry 1996; 168:607-11.
14. Watt S, Sword W, Krueger P, Sheehan D; Ontario Mother & Infant Survey. A cross-sectional study of early identification of postpartum depression: implications for primary care providers from The Ontario Mother & Infant Survey. BMC Fam Pract 2002; 11:3-5.
15. Uwakwe R, Okonkwo JE. Affective (depressive) morbidity in puerperal Nigerian women: validation of the Edinburgh Postnatal Depression Scale. Acta Psychiatr Scand 2003; 107:251-9.
16. Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression: development of the 10-item Edinburgh Postnatal Depression Scale. Br J Psychiatry 1987; 150:782-6.
17. Barros FC, Victora CG, Barros AJD, Santos IS, Albernaz E, Matijasevich A, et al. The challenge of reducing neonatal mortality in middle-income countries: findings from three Brazilian birth cohorts in 1982, 1993, and 2004. Lancet 2005; 365:847-54.
18. Da Silva VA, Moraes-Santos AR, Carvalho MS, Martins MLP, Teixeira NA. Prenatal and postnatal depression among low-income Brazilian women. Braz J Med Biol Res 1998; 31:799-804.
19. Organização Mundial da Saúde. Classificação de Transtornos Mentais e de Comportamento da CID-10. Critérios diagnósticos para pesquisa. Porto Alegre: Editora Artes Médicas; 1998.
20. Eberhard-Gran M, Eskild A, Tambs K, Opjordsmoen S, Samuelsen SO. Review of validation studies of the Edinburgh Postnatal Depression Scale. Acta Psychiatr Scand 2001; 104:243-9.
21. Santos MFS, Martins FC, Pasquali L. Escala de auto-avaliação de depressão pós-parto: estudo no Brasil. Rev Psiquiatr Clin 1999; 26:90-5.
22. Fletcher RH, Fletcher SW, Wagner EH. Epidemiologia clínica: elementos essenciais. Porto Alegre: Editora Artes Médicas; 1996.
23. Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JHP, et al. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA 1999; 282:1061-6.
24. Rutjes AW, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. CMAJ 2006; 174:469-76.
25. World Health Organization. Postpartum care of the mother and newborn: a practical guide. http://www.who.int/reproductive-health/publications/msm_98_3/msm_98_3_14.html (accessed on 29/Mar/2007).
26. Blenning CE, Paladine H. An approach to the postpartum office visit. Am Fam Physician 2005; 72:2491-6.

Submitted on 04/May/2006
Final version resubmitted on 07/Feb/2007
Approved on 14/Feb/2007