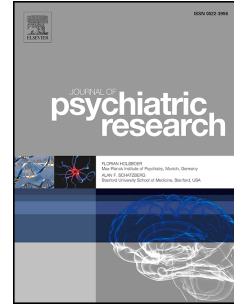


Journal Pre-proof

The diagnostic accuracy of the Edinburgh Postnatal Depression Scale without the self-harm item: Does culture matter?

Alberto Stefana, Loredana Cena, Alice Trainini, Gabriella Palumbo, Antonella Gigantesco, Fiorino Mirabella



PII: S0022-3956(24)00277-2

DOI: <https://doi.org/10.1016/j.jpsychires.2024.05.018>

Reference: PIAT 6229

To appear in: *Journal of Psychiatric Research*

Received Date: 31 July 2023

Revised Date: 31 January 2024

Accepted Date: 3 May 2024

Please cite this article as: Stefana A, Cena L, Alice T, Palumbo G, Gigantesco A, Mirabella F, The diagnostic accuracy of the Edinburgh Postnatal Depression Scale without the self-harm item: Does culture matter?, *Journal of Psychiatric Research*, <https://doi.org/10.1016/j.jpsychires.2024.05.018>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2024 Published by Elsevier Ltd.

Title: The diagnostic accuracy of the Edinburgh Postnatal Depression Scale without the self-harm item: Does culture matter?

Author names and affiliations: Alberto Stefana,^a Loredana Cena,^b Alice Trainini,^b Gabriella Palumbo,^c Antonella Gigantesco,^{c*} & Fiorino Mirabella^{c*}

^a Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy.

^b Department of Clinical and Experimental Sciences, Section of Neuroscience, Observatory of Perinatal Clinical Psychology, University of Brescia, Brescia, Italy.

^c Center for Behavioural Sciences and Mental Health, National Institute of Health, Rome, Italy.

* These authors equally contributed to this work and should be considered co-last authors.

Correspondence to: Dr Alberto Stefana, PhD, Department of Brain and Behavioral Sciences, University of Pavia, Via Forlanini 6, 27100 Pavia, Italy. E-mail: alberto.stefanai@unipv.it

Declaration of interest: The authors have no conflicts of interest to declare. All co-authors have seen and agree with the contents of the manuscript and there is no financial interest to report.

The first author (AS) has worked on this article thanks to funds from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 101030608.

Article data: number of tables: 1; article length: 1118 (excluding references) words.

The diagnostic accuracy of the Edinburgh Postnatal Depression Scale without the self-harm item: Does culture matter?

To the Editor:

We read with keen interest the recent article by Chen et al. (2023), in which the authors evaluated the performance of the Edinburgh Postnatal Depression Scale (EPDS) without the self-harm item, called EPDS-9, compared to the complete EPDS, called EPDS-10. They focused on identifying depression among people who are pregnant or postpartum. The authors concluded that the shortened EPDS-9 performs as well as the EPDS-10, suggesting it as a potential replacement for the full-length EPDS.

Our research partially supports the findings of Chen et al. (2023). Our study sample comprises 1153 pregnant women and 309 postpartum women. These participants were enrolled from 11 healthcare centers located throughout Italy (*masked citation*). The characteristics of the participants are detailed in a separate publication (*masked citation*). Trained psychologists used unstructured clinical interviews and patient-rated Patient Health Questionnaire-9 (PHQ-9) and EPDS questionnaires to evaluate participants' depression.

Our findings indicate a correlation of .998 between EPDS-9 and EPDS-10, observed in both the antepartum and postpartum groups. Only 1% of the participants were negative at EPDS-9 cutoff points of <10 but had a non-zero EPDS item 10 score, and 2% at EPDS-9 cutoff points of <13. Furthermore, EPDS-9 demonstrated excellent accuracy in distinguishing EPDS-10-based depression screening in both perinatal groups, in each of the four commonly used cutoff scores (Levi et al., 2020; Quip et al., 2023).

We used the PHQ-9 as a criterion to compare the performance of the EPDS-9 *versus* EPDS-10, using a cut-off value of 13 (which is indicated as the most appropriate for the detection of major depression in perinatal people [Levi et al., 2019]). EPDS-9 and EPDS-10 demonstrated comparable sensitivity, specificity, and area under the curve (AUC) performances. In the antepartum group, both the EPDS-9 and EPDS-10 (a) show declining sensitivity with increasing cutoff values, (b) have high specificity across all cutoff values, and (c) have AUC values that suggest they perform reasonably well,

though their performance declines with increasing cutoff values. Comparison of AUC values between EPDS-9 and EPDS-10 suggests that there are no significant differences in performance between the two versions of EPDS at cutoff values of 10, 11, and 13. However, there appears to be a significant difference in performance at a cutoff value of 12, with the EPDS-10 performing better. Regarding the postpartum group, although the AUC remains relatively high for both EPDS-9 and EPDS-10 across all cutoff values, the equivalence tests showed a statistically significant difference at all cutoff values (see Table 1), indicating that there is a significant difference in overall test performance. Specifically, the EPDS-10 outperforms the EPDS-9 at all cutoff values.

We also examined the predictive potential of EPDS-9 for responses to the EPDS self-harm item (item 10). The AUC of EPDS-9 against self-harm responses varied depending on the frequency level, which could be an area for further study. Specifically, EPDS-9's AUC against self-harm above the frequency of "hardly" ranged from 0.716 to 0.826, except for cutoff 13 in the antepartum group, where it dropped significantly to 0.288. This decrease in AUC at the cut-off point of 13 suggests that EPDS-9's ability to predict self-harm responses decreases when this more conservative threshold is used. The AUC against self-harm above the frequency of "sometimes" and "often" ranged, respectively, from 0.712 to 0.826 and from 0.445 to 0.675. These variations emphasize the importance of considering frequency when examining self-harm predictions. Table 1 shows the sensitivity, specificity, and AUC for each cutoff value.

Based on our study, we propose two main findings that support those of Chen et al. First, EPDS-10 and EPDS-9 are strongly correlated. Second, EPDS-9 exhibited similar sensitivity and specificity in screening major depression among pregnant and postpartum women, compared to full EPDS, across the most commonly used cutoff points.

However, unlike the Japanese sample of Chen et al., EPDS-9 did not predict the responses of Italian participants to the self-harm item as accurately. We found this discrepancy when comparing the differentiation performance of EPDS-9 *versus* EPDS-10 using the PHQ-9 as a criterion. Likely, the discrepancy is due to the use of different instruments although as Kessler Psychological Distress Scale (K6) (used by Chen et al.) and PHQ-9 showed a strong correlation (Cotton et al., 2021).

It is here important to remember that the EPDS was originally developed in English (Cox et al., 1987). Consequently, both our study and that of Chen et al. employed translated versions of the scale. Although both the Japanese and the Italian translations have been validated (Benvenuti et al., 1999; Okano et al., 1996) and shown to be reliable and valid measures for perinatal depression (Kubota et al., 2018; Stefana et al., 2023) and have demonstrated a similar factor structure (which includes aspects of anxiety and anhedonia (Kubota et al., 2014; Mirabella et al., 2024), the translation process may contribute to some of the inconsistencies in the data. This highlights a critical issue: the necessity of establishing cross-cultural validity for psychological inventories.

Cultural variations in the subjective experience and expression of affective disorders must be taken into account in clinical assessment (Kiermaier & Groleau, 2001). They may significantly shape the manifestation of depression symptomatology and impact the openness to answer questions about self-harm, as suggested by numerous studies. Mental health issues such as depression can present differently in various cultures due to differences in social norms, belief systems, and levels of stigma associated with mental health (Kleinman & Good, 1985). In some societies, such as the Chinese one, psychological symptoms may be expressed more somatically, which may influence the detection of depressive symptoms through tools such as the EPDS (Ryder et al., 2008).

Concerning self-harm and suicidal ideation, cultural factors can significantly influence the willingness to disclose such experiences. For example, some cultures may have high levels of stigma associated with mental health conditions or self-harm behaviors, making individuals less likely to report these experiences openly (Chu et al., 2010). Additionally, cultures like, for example, the Chinese one prioritize collective identity over individualism may see a higher level of self-stigma, resulting in a lower level of openness about mental health struggles, including self-harm (Yang et al., 2007).

Therefore, it is crucial to keep cultural factors in mind when interpreting the effectiveness of measures such as EPDS-9 and EPDS-10 in different cultures and perinatal populations (pregnant *versus* postpartum people). The variance between Chen et al.'s and our samples in terms of the predictive precision of EPDS-9 for self-harm responses underscores the need for culturally sensitive approaches in the detection of depression. More research is needed to understand the specific cultural

factors at play in the various phases of the perinatal process and adapt the instruments accordingly to improve their validity and reliability.

Lastly, Chen et al.'s suggestion to omit the self-harm item in order to help avoid confusion and potential psychological distress brought to the responders should be considered with caution. Overreliance on item 10 can surely lead to a strain on resources due to mandatory follow-up assessments, but when psychological assessment is done well it is always therapeutic to some degree. Furthermore, as we explained before, in certain cultures (and more generally, in certain people) a 0 score on item 10 does not mean a 0 risk of suicide.

In conclusion, although EPDS-9 shows a performance similar to that of EPDS-10 in the screening of major depression, we recommend the use of the full EPDS. The variance in predictive accuracy between different population samples highlights the need for future research to further validate EPDS-9 in specific cultures and perinatal populations.

References

- Benvenuti, P., Ferrara, M., Niccolai, C., Valoriani, V., & Cox, J. L. (1999). The Edinburgh Postnatal Depression Scale: validation for an Italian sample. *Journal of Affective Disorders*, *53*(2), 137–141.
- Chen, C., Okubo, R., Okawa, S., Nakagawa, S., & Tabuchi, T. (2023). The diagnostic accuracy of the Edinburgh Postnatal Depression Scale without the self-harm item. *Journal of Psychiatric Research*, *165*, 70–76.
- Chu, J. P., Goldblum, P., Floyd, R., & Bongar, B. (2010). The cultural theory and model of suicide. *Applied and Preventive Psychology*, *14*, 25-40.
- Cotton, S. M., Menssink, J., Filia, K., et al. (2021). The psychometric characteristics of the Kessler Psychological Distress Scale (K6) in help-seeking youth: What do you miss when using it as an outcome measure?. *Psychiatry Research*, *305*, 114182.
- Cox, J. L., Holden, J. M., & Sagovsky, R. (1987). Detection of postnatal depression: development of the 10-item Edinburgh Postnatal Depression Scale. *British Journal of Psychiatry*, *150*(6), 782-786.
- Kim, J. J., La Porte, L. M., Saleh, M. P., Allweiss, S., Adams, M. G., Zhou, Y., & Silver, R. K. (2015). Suicide risk among perinatal women who report thoughts of self-harm on depression screens. *Obstetrics and Gynecology*, *125*, 885–893.
- Kirmayer, L. J., & Groleau, D. (2001). Affective disorders in cultural context. *The Psychiatric Clinics of North America*, *24*, 465–vii.
- Kleinman, A., & Good, B. (Eds.). (1985). *Culture and depression: Studies in the anthropology and cross-cultural psychiatry of affect and disorder*. University of California Press.
- Kubota, C., Inada, T., Nakamura, Y., Shiino, T., Ando, M., Aleksic, B., ... & Ozaki, N. (2018). Stable factor structure of the Edinburgh Postnatal Depression Scale during the whole peripartum period: Results from a Japanese prospective cohort study. *Scientific Reports*, *8*(1), 17659.
- Levis, B., Benedetti, A., Thombs, B. D., & DEPRESSion Screening Data (DEPRESSD) Collaboration (2019). Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. *BMJ*, *365*, l1476.
- Levis, B., Negeri, Z., Sun, Y., Benedetti, A., Thombs, B. D., & DEPRESSion Screening Data (DEPRESSD) EPDS Group. (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.
- Mirabella, F., Gigantesco, A., Palumbo, G., Cena, L., Trainini, A., & Stefana, A., (2024). Screening for prenatal and post, natal maternal depression: comparative performance of the Edinburgh Postnatal Depression Scale and Patient Health Questionnaire-9. *Annali dell'Istituto Superiore di Sanità*.

- Okano, T. M. M., Masuji, F., Tamaki, R., Nomura, J., & Miyaoko, H. (1996). Validation and reliability of Japanese version of the EPDS. *Archives of Psychiatric Diagnostics and Clinical Evaluation*, 7, 525-533
- Qiu, X., Wu, Y., Sun, Y. et al. (2023) Individual participant data meta-analysis to compare EPDS accuracy to detect major depression with and without the self-harm item. *Scientific Reports*, 13, 4026.
- Ryder, A. G., Yang, J., Zhu, X., et al. (2008). The cultural shaping of depression: somatic symptoms in China, psychological symptoms in North America? *Journal of Abnormal Psychology*, 117, 300-313.
- Stefana, A., Langfus, J. A., Palumbo, G., Cena, L., Trainini, A., Gigantesco, A., & Mirabella, F. (2023). Comparing the factor structures and reliabilities of the EPDS and the PHQ-9 for screening antepartum and postpartum depression: a multigroup confirmatory factor analysis. *Archives of Women's Mental Health*, 26(5), 659–668.
- Yang, L. H., Kleinman, A., Link, B. G., Phelan, J. C., Lee, S., & Good, B. (2007). Culture and stigma: Adding moral experience to stigma theory. *Social Science & Medicine*, 64, 1524-1535.

Table 1
Sensitivity, specificity, and AUC values

Performance of EPDS-9 against EPDS-10-based screening of depression				
	Cutoff = 10	Cutoff = 11	Cutoff = 12	Cutoff = 13
Antepartum	Sensitivity = 0.962 Specificity = 0.997 AUC = 0.979	Sensitivity = 0.968 Specificity = 1 AUC = 0.985	Sensitivity = 0.968 Specificity = 1 AUC = 0.984	Sensitivity = 0.979 Specificity = 1 AUC = 0.990
Postpartum	Sensitivity = 1 Specificity = 0.996 AUC = 0.998	Sensitivity = 0.971 Specificity = 0.998 AUC = 0.984	Sensitivity = 0.946 Specificity = 1 AUC = 0.973	Sensitivity = 0.961 Specificity = 1 AUC = 0.980
Comparison of performance between EPDS-9 and EPDS-10 against PHQ-9-based screening of depression.				
	Cutoff = 10	Cutoff = 11	Cutoff = 12	Cutoff = 13
Antepartum	EPDS-10 Sensitivity = 0.651 Specificity = 0.929 AUC = 0.790	EPDS-10 Sensitivity = 0.602 Specificity = 0.950 AUC = 0.776	EPDS-10 Sensitivity = 0.494 Specificity = 0.969 AUC = 0.732	EPDS-10 Sensitivity = 0.410 Specificity = 0.984 AUC = 0.697
	EPDS-9 Sensitivity = 0.651 Specificity = 0.931 AUC = 0.791 AUC difference = 0.001 $p = 0.045$ Equivalent = true	EPDS-9 Sensitivity = 0.602 Specificity = 0.951 AUC = 0.777 AUC difference = 0.001 $p = 0.051$ Equivalent = true	EPDS-9 Sensitivity = 0.470 Specificity = 0.971 AUC = 0.720 AUC difference = 0.012 $p < 0.001$ Equivalent = false	EPDS-9 Sensitivity = 0.041 Specificity = 0.986 AUC = 0.698 AUC difference = 0.001 $p = 0.076$ Equivalent = true
Postpartum	EPDS-10 Sensitivity = 1 Specificity = 0.851 AUC = 0.925	EPDS-10 Sensitivity = 0.894 Specificity = 0.874 AUC = 0.884	EPDS-10 Sensitivity = 0.851 Specificity = 0.916 AUC = 0.884	EPDS-10 Sensitivity = 0.787 Specificity = 0.958 AUC = 0.873
	EPDS-9 Sensitivity = 0.936 Specificity = 0.840 AUC = 0.888 AUC difference = 0.037 $p < 0.001$ Equivalent = false	EPDS-9 Sensitivity = 0.872 Specificity = 0.874 AUC = 0.873 AUC difference = 0.011 $p < 0.001$ Equivalent = false	EPDS-9 Sensitivity = 0.830 Specificity = 0.920 AUC = 0.875 AUC difference = 0.009 $p < 0.001$ Equivalent = false	EPDS-9 Sensitivity = 0.766 Specificity = 0.958 AUC = 0.862 AUC difference = 0.011 $p < 0.001$ Equivalent = false
Performance of EPDS-9 against thoughts of self-harm.				
	EPDS-9 cutoff = 10	EPDS-9 cutoff = 11	EPDS-9 cutoff = 12	EPDS-9 cutoff = 13
Antepartum	\geq hardly ever Sensitivity = 0.106 Specificity = 0.462 AUC = 0.716	\geq hardly ever Sensitivity = 0.083 Specificity = 0.462 AUC = 0.728	\geq hardly ever Sensitivity = .055 Specificity = 0.462 AUC = 0.742	\geq hardly ever Sensitivity = .0380 Specificity = 0.538 AUC = 0.288
	\geq sometimes Sensitivity = 0.538 Specificity = 0.894 AUC = 0.716	\geq sometimes Sensitivity = 0.538 Specificity = 0.917 AUC = 0.728	\geq sometimes Sensitivity = 0.538 Specificity = 0.945 AUC = 0.742	\geq sometimes Sensitivity = 0.462 Specificity = 0.962 AUC = 0.712
	\geq often Sensitivity = 0.000 Specificity = 0.889 AUC = 0.445	\geq often Sensitivity = 0.000 Specificity = 0.912 AUC = 0.456	\geq often Sensitivity = 0.000 Specificity = 0.939 AUC = 0.470	\geq often Sensitivity = 0.000 Specificity = 0.958 AUC = 0.479
Postpartum	\geq hardly ever Sensitivity = 0.266 Specificity = 0.125 AUC = 0.805	\geq hardly ever Sensitivity = 0.223 Specificity = 0.125 AUC = 0.826	\geq hardly ever Sensitivity = 0.183 Specificity = 0.375 AUC = 0.721	\geq hardly ever Sensitivity = 0.014 Specificity = 0.375 AUC = 0.743
	\geq sometimes Sensitivity = 0.875 Specificity = 0.734 AUC = 0.805	\geq sometimes Sensitivity = 0.875 Specificity = 0.777 AUC = 0.826	\geq sometimes Sensitivity = 0.625 Specificity = 0.817 AUC = 0.721	\geq sometimes Sensitivity = 0.625 Specificity = 0.860 AUC = 0.743
	\geq often Sensitivity = 0.500 Specificity = 0.720 AUC = 0.610	\geq often Sensitivity = 0.500 Specificity = 0.762 AUC = 0.631	\geq often Sensitivity = 0.500 Specificity = 0.808 AUC = 0.654	\geq often Sensitivity = 0.500 Specificity = 0.850 AUC = 0.675

Title: The diagnostic accuracy of the Edinburgh Postnatal Depression Scale without the self-harm item: Does culture matter?

Author names and affiliations: Alberto Stefana,^a Loredana Cena,^b Alice Trainini,^b Gabriella Palumbo,^c Antonella Gigantesco,^{c*} & Fiorino Mirabella^{c*}

^a Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy.

^b Department of Clinical and Experimental Sciences, Section of Neuroscience, Observatory of Perinatal Clinical Psychology, University of Brescia, Brescia, Italy.

^c Center for Behavioural Sciences and Mental Health, National Institute of Health, Rome, Italy.

Declaration of interest: The authors have no conflicts of interest to declare. All co-authors have seen and agree with the contents of the manuscript and there is no financial interest to report.