Short communication: Detecting depression after pregnancy: the validity of the K10 and K6 in Burkina Faso

R. F. Baggaley1, R. Ganaba2, V. Filippi1, M. Kere3, T. Marshall1, I. Sombie2, K. T. Storeng1 and V. Patel1

1 Department of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, UK
2 Centre Muraz, Bobo-Dioulasso, Burkina Faso
3 Centre hospitalier universitaire Souro Sanou, Bobo-Dioulasso, Burkina Faso

Introduction

Depression is one of the world’s most disabling illnesses in both developed and developing countries. It contributes to mortality through suicide, injury, cardiovascular disease and other health problems (Mathers & Loncar 2006) and interferes with the ability to work or function in non-income roles (World Health Organisation 2001). It affects healthy lifestyles, risk of HIV infection, treatment adherence and disease prognosis (Catalan 1999), as well as cohesive social function, which impacts on communities’ economies (Islam et al. 2006; Gureje & Jenkins 2007). Postnatal depression is common, with prevalence estimates of 10–15% in developed countries (O’Hara & Swain 1996; Beck & Gable 2001) and 5–60% in developing countries (Lee et al. 1998; Cooper et al. 1999; Affonso et al. 2000; Chandran et al. 2002; Patel et al. 2002; Rahman et al. 2003), indicating the extent of the problem, the considerable differences in risk between settings, and difficulties in measurement. Postnatal depression has serious consequences for maternal physical health (Cox et al. 1982; Cooper & Murray 1998), and is associated with adverse birth and child outcomes (Patel et al. 2004).

As awareness of depression in developing countries has increased, prevalence of depression and changes in prevalence over time need to be monitored. As gold standard tools for detecting depression, such as the Composite International Diagnostic Interview, are not feasible for large-scale epidemiological studies in developing countries due to the time and professional expertise required, easily administered and quick screening tools are required.

A number of screening questionnaires for depression exist, including the General Health Questionnaire, the Self Reporting Questionnaire, the Edinburgh Postnatal Depression Scale (EPDS) and the K10. The K10 is a short questionnaire developed to screen for depression, with or without anxiety, by determining a composite score based on participants’ responses (Kessler et al. 2002). It is ideal for monitoring prevalence of depressive disorders on a large scale. The K10 has been derived on the basis of a systematic
process, extensively tested and validated in various regions (Andrews & Slade 2001; Kessler et al. 2002; Furukawa et al. 2003). It has strong psychometric properties and excellent ability to distinguish cases from non-cases. Its wording is simple, with short questions and clear response categories. Furthermore, it is the instrument being used in the World Mental Health Surveys, which include centres in West Africa, and it has recently been validated for use in Nigeria (Gureje et al. 2006). Despite this, relatively few studies have been conducted evaluating its use in Africa and there is currently no threshold score for determining detection of depression for use in the field in these settings. The K10 has not previously been validated as a tool for use in Francophone Africa or in predicting postnatal depression in any setting. The objective of this sub-study was to investigate the validity of the K10 and K6 in Burkina Faso for the detection of postnatal depression.

Methods

K10 and K6 scores

The K10 is a 10-item scale with five response categories ranked on a five-point scale, with the score being the sum of these responses (details in the Appendix). The K6 consists of a subset of six of these items. Scores are 0–40 for the K10 and 0–24 for the K6. The WHO-issued English language version of the K10 questionnaire was translated into West African French and the local languages of Mooré and Dioula using the standardised WHO translation and back-translation protocol (WHO) and pre-tested before being finalised (http://www.who.int/substance_abuse/research_tools/translation/en/index.html).

The K10 questionnaire was administered within a larger cohort study of postpartum women and women with early pregnancy loss (miscarriage, termination and ectopic pregnancy) by trained interviewers at 3, 6 and 12 months post-pregnancy (Filippi et al. in press). Interviewers took a 1-day training course with a local psychiatrist on the rationale and methods for the K10.

Clinical diagnosis

Of the cohort participants, 61 additionally completed a separate, diagnostic interview with a local psychiatrist, to provide the ‘gold standard’ for assessment of depression against which to compare the K10. The psychiatrist was blind to the subjects’ K10 scores. Women were selected in an attempt to over-sample from those with higher K10 scores in their most recent interview to gain a larger sample of probable cases of depression, but otherwise were chosen at random. Diagnoses were based on the ICD-10 criteria for Mental and Behavioural Disorders (http://www.who.int/classifications/icd/en/GRNBOOK.pdf), using a checklist of factors for consideration to facilitate standardisation of diagnoses. Women were classified as cases if diagnosed with depression, regardless of severity. Any woman diagnosed as depressed was given a follow-up appointment with the psychiatrist for further evaluation and treatment. The interview was taken within 3 days of a K10 assessment, at either the 3 month (n = 29, 48%) or 6 month (n = 32, 52%) postpartum interview.

Statistical analysis

Data were analysed using stata version 9. Cronbach’s alpha tests internal consistency by assessing how well a set of items measures a single, one-dimensional outcome by determining their correlation (Cronbach 1951). Alpha ≥0.70 was considered satisfactory (Nunnally & Bernstein 1994). Reliability of each item was also assessed by measuring the correlation between each item and the overall K10 and K6 scores using the Pearson product-moment correlation. To test whether the K10 and K6 are valid indices predicting risk of depression, the Kruskal–Wallis test was used to evaluate whether K10 and K6 scores were significantly different for cases and non-cases. Receiver Operating Characteristic (ROC) curves were constructed to investigate optimum cut-off scores for identifying cases.

Ethical approval

Ethical approval was granted by the ethical committees of the London School of Hygiene and Tropical Medicine and Centre Muraz, Bobo-Dioulasso, Burkina Faso. Informed consent was obtained from all informants.

Results

All 61 respondents completed all items of the K10 questionnaire. The age range was 17–46 years, mean 26 (SD = 7) years; 39 (64%) were married, and 28 (46%) had completed six or more years of education. The mean K10 and K6 scores were 10.7 and 7.0, respectively, while the clinical assessment found that 27 (44%) women were cases.

Reliability

Cronbach’s alpha coefficient was 0.87 and 0.78 for K10 and K6 scores, respectively, indicating satisfactory reliability. Item-total correlation, assessed using the Pearson product-moment statistic, varied from 0.44 to 0.83 for the
K10 scale and from 0.44 to 0.81 for the K6 scale, and all correlations were highly significant (\(P = 0.0005\) or lower, data not shown), indicating satisfactory reliability of all items.

**Validity**

Cases had significantly higher K10 scores than non-cases [mean 14.3 (SD = 6.9) and 7.8 (SD = 6.2), respectively (\(P = 0.0003\))]$. Results were similar for the K6 score [mean 9.3 (SD = 4.4) and 5.3 (SD = 3.8), respectively (\(P = 0.0008\))]. ROC curves plotting the relationship between each score and the clinical diagnosis of depression were not significantly different for the K10 and K6 (\(P = 0.239\)), with areas under the curve of 0.77 and 0.75, respectively (Figure 1). A \(\geq 6\) cut-off on the K10 score, the equivalent of which is used elsewhere to predict moderate risk of depression (Andrews & Slade 2001), gave sensitivity 85% and specificity 41%. The optimum cut-off for defining cases of depression for the K10 score was \(\geq 12\) according to the ROC curve, giving the optimum values of sensitivity (74%) and specificity (76%), but \(\geq 14\) gave the greatest percentage correctly classified (77% for \(\geq 14\) compared with 75% for \(\geq 12\)). The \(\geq 14\) cut-off gave sensitivity 59% and specificity 91%. For the K6, cut-offs of \(\geq 9\) or \(\geq 10\) could have been used because no respondent had a K6 of 9 and so all outcomes were identical for the two cut-offs. Cut-offs of \(\geq 9\), \(\geq 10\) and \(\geq 11\) each give the highest percentage correctly classified (74%), but \(\geq 9\) and \(\geq 10\) perform better in the ROC analysis.

**Discussion**

This small-scale validation study suggests that the West African French and local language versions of the K10 and K6 questionnaires are reasonably valid indicators of depression among postpartum women in Burkina Faso. It indicates higher K10 and K6 cut-off scores for depression for this setting compared to elsewhere (Andrews & Slade 2001). Such differences may be due to variation in the discriminatory power of different items in the K10 and K6 scores or differences in baseline prevalence of depression across settings, among other reasons (Goldberg et al. 1998).

While a cut-off of 12 for the K10 performed better in the ROC analysis, this was calculated for a group oversampling for depression and therefore an artificially high prevalence of depression. A 14 cut-off may be preferable in this setting; for service provision where resources are scarce, positive predictive value is often a crucial factor. Similarly, a 10 cut-off for the K6 is recommended. The K10 measures risk of depression rather than attempting to diagnose it, and if it is to be used to identify women at risk for depression for referral for further evaluation, a lower cut-off would be recommended. In this particular context, the tool is used to monitor prevalence of depression and is thus used on a large sample; therefore the brevity of the questionnaire and achieving higher levels of sensitivity and specificity need to be weighed against each other.

These cut-offs provide categorical assessment of depression (dividing samples into those at high and low risk of depression), but dimensional assessments based on mean K10 and K6 scores are also recommended for epidemiological and sociological research, as they retain greater statistical power (Kessler 2002). The use of such an easily administered tool for mental health research will play a vital role in raising awareness of the extent and impact of depression within all resource-poor settings.

**Acknowledgements**

We thank the entire Immpact team in Burkina Faso, including all the interviewers, and particularly Thomas Ouédraogo and Nicolas Méda, as well as all the women who participated in the study. This work was undertaken...
as part of an international research programme - Immpact (http://www.abdn.ac.uk/immpact), funded by the Bill & Melinda Gates Foundation, the Department for International Development, the European Commission and USAID. The funders have no responsibility for the information provided or views expressed in this paper. The views expressed herein are solely those of the authors.

**Appendix 1**

**K10 and K6 score items**

Questions from the English language version of the K10 and K6, upon which the West African French translation was based: 

In the last 30 days, about how often did you feel... 

(a) ...tired out for no good reason? 
(b) ...nervous? \(^1\) (if never go to question d) 
(c) ...so nervous that nothing could calm you down? 
(d) ...hopeless? \(^1\) 
(e) ...restless or fidgety? \(^1\) (if never go to question g) 
(f) ...so restless that you could not sit still? 
(g) ...depressed? (if never go to question i) 
(h) ...so depressed that nothing could cheer you up? \(^1\) 
(i) ...that everything was an effort? \(^1\) 
(j) ...worthless? \(^1\)

\(^1\)Denotes questions used for the K6 score.

\(^2\)Replies of more than 'never' i.e. a score more than 0 for questions c, f or h would be dependent on a reply of more than 'never' for the preceding question in each case. Therefore if a respondent replied 'never' to question b (or question e or g) then they skipped question c (or question f or h), which was automatically coded as 0 i.e. never. This again saves time when administering the questionnaire.

Scoring of answers: 

Never = 0 
Occasionally = 1 
Sometimes = 2 
Most of the time = 3 
All of the time = 4 

The total K10 and K6 scores are computed using the equation:

\[
\text{Total score} = \frac{\text{Sum of items scores}}{\text{Number of valid items}} \times \text{Number of items,}
\]

with the result being rounded to the nearest whole number. If any required item has not been completed, it is excluded from the calculation and not counted as a valid item. Number of items is 10 for the K10 and six for the K6. If more than one of the items constituting the total score is not valid then the total score is set as missing.

**References**


Detecting depression after pregnancy


### Corresponding Author

R.F. Baggaley, Department of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, Keppel Street, London WC1E 7HT, UK. Tel.: 02079272469; Fax: 02076374314; E-mail: rebecca.baggaley@lshtm.ac.uk

---

Détection de la dépression après grossesse: la validité du K10 et du K6 au Burkina-Faso

**Objetivo** Las K10 y los K6 son escala de clasificación corta diseñadas para detectar individuos a riesgo de depresión con o sin ansiedad. Malgré leur emploi étendu, elles n’ont pas été encore validées pour la détection de la dépression postnatale. Nous décrivons ici la validité de ces échelles pour la détection de la dépression postnatale au Burkina-Faso.

**Método** La versión inglesa del cuestionario K10 a sido traducida en las lenguas locales de África Occidental y francés para su uso en Burkina Faso. Se compilaron los puntajes de 61 mujeres con la entrevista de diagnóstico realizada por un psiquiatra local dentro de los 3 días de administración del K10.

**Resultados** La evaluación clínica encontró que 27 mujeres (44%) eran casos probables de depresión. La consistencia interna de los puntajes del K10 y K6, definidas por el coeficiente alpha de Cronbach, era 0.87 y 0.78 respectivamente, indicando una fiabilidad satisfactoria. La sensibilidad y especificidad para el K10 y K6, respectivamente, para identificar a las mujeres a alto riesgo de depresión son del 59% y 85%.

**Conclusión** Este estudio sugiere que el K10 y K6 son medidas razonablemente válidas para la depresión en mujeres que han acogido en Burkina Faso y pueden utilizarse como herramientas relativamente baratas para estimar la prevalencia de esta condición en países en vías de desarrollo.

**Mots clés** dépression post-partum, diagnostic, échelle d’évaluation K10, échelle d’évaluation K6, validité, Burkina-Faso

Detectando la depresión después del embarazo: la validez del K10 y K6 en Burkina Faso

**Objetivo** K10 y K6 son escalas de clasificación corta diseñadas para detectar individuos a riesgo de desórdenes depresivos, con o sin ansiedad.

**Método** Se tradujo la versión inglesa del cuestionario del K10 al francés y a las lenguas locales de África Occidental, para su uso en Burkina Faso. Se compararon los puntajes de 61 mujeres con la entrevista de diagnóstico realizada por el psiquiatra local dentro de los 3 días de administración del K10.

**Resultados** La evaluación clínica encontró que 27 mujeres (44%) eran casos probables de depresión. La consistencia interna de los puntajes del K10 y K6, definidos por el coeficiente alfa de Cronbach, era 0.87 y 0.78 respectivamente, indicando una fiabilidad satisfactoria. El desempeño de los puntajes no era significativamente diferente, con áreas bajo la curva de 0.77 y 0.75 para el K10 y K6 para identificar a las mujeres a alto riesgo de depresión. Para estimar la prevalencia de depresión, hemos sugerido puntos de corte de 14 para el K10 y entre 9 y 11 para el K6, con el fin de identificar mujeres con alto riesgo de depresión. A ≥14, el K10 tiene un 59% de sensibilidad, 91% de especificidad; a ≥10, el K6 tiene un 59% de sensibilidad y un 85% de especificidad.

**Conclusión** Este estudio sugiere que el K10 y K6 son medidas razonables para la depresión posparto en mujeres de Burkina Faso, y pueden utilizarse como herramientas relativamente baratas para estimar la prevalencia de esta condición en países en vías de desarrollo.

**Palabras clave** depresión posparto, diagnóstico, escala de clasificación K10, escala de clasificación K6, validez, Burkina Faso