Validation and factor structure of the Thai version of the EURO-D scale for depression among older psychiatric patients

Abbreviated title: Validation of Euro-D, Thai version

List of authors
Tawanchai Jirapramukpitak¹*
Niphon Darawuttimapraporn²
Sureeporn Punpuing²
Melanie Abas³

¹Department of Psychiatry, Faculty of Medicine, Thammasat University, Klong Luang, Pathumthani, Thailand
²Institute for Population and Social Research, Mahidol University, Salaya, Nakhonpathom, Thailand
³Institute of Psychiatry, King’s College London, Denmark Hill, London, UK

*Correspondence to: Dr T. Jirapramukpitak, Department of Psychiatry, Faculty of Medicine, Thammasat University, Paholyothin Rd., Klong Luang, Pathumthani, Thailand, 12120.
Tel: 66 (0)2 926 9757-8, Fax: 66 (0)2 926 9755, Email: tawanchaij@gmail.com
Validation and factor structure of the Thai version of the EURO-D scale for depression among older psychiatric patients

Name of journal: Aging & Mental Health

Objectives: To assess the concurrent and the construct validity of the Euro-D in older Thai persons.

Method: Eight local psychiatrists used the major depressive episode section of the Mini International Neuropsychiatric Interview to interview 150 consecutive psychiatric clinic attendees. A trained interviewer administered the Euro-D. We used receiver operating characteristic (ROC) analysis to assess the overall discriminability of the Euro-D scale and principal components factor analysis to assess its construct validity.

Results: The area under the ROC curve for the Euro-D with respect to major depressive episode was 0.78 (95% CI 0.70-0.90) indicating moderately good discriminability. At a cut-point of 5/6 the sensitivity for major depressive episodes is 84.3%, specificity 58.6%, and kappa 0.37 (95% CI 0.22-0.52) indicating fair concordance. However at the 3/4 cut-point recommended from European studies there is high sensitivity (94%) but poor specificity (34%). The principal components analysis suggested four factors. The first two factors conformed to affective suffering (depression, suicidality and tearfulness) and motivation (interest, concentration and enjoyment). Sleep and appetite constituted a separate factor whereas pessimism loaded on its own factor.

Conclusion: Among Thai psychiatric clinic attendees Euro-D is moderately valid for major depression. A much higher cut-point may be required than that which is usually advocated. The Thai version also shares two common factors as reported from most of previous studies.

Keywords: aged; depressive disorders; Thailand; validation studies
Validation and factor structure of the Thai version of the EURO-D scale for depression among older psychiatric patients

Introduction

Depression in older persons is reported to be associated with substantially reduced quality of life and increased mortality (Penninx, Geerlings et al., 1999; Penninx, Leveille, Ferrucci, van Eijk, & Guralnik, 1999) and increased use of all health and social service resources (Watts et al., 2002). In spite of its public health significance, levels of recognition and treatment of depression by physician are reported to be rather low (Crawford, Prince, Menezes, & Mann, 1998; Dearman, Waheed, Nathoo, & Baldwin, 2006; Koenig, 2007). There has been little research from Thailand on depression in older people. Given the importance of detection of depression in older people, there is a need to validate a Thai version of a depression screening instrument.

Euro-D was an instrument developed by 14 European countries for screening depression in the elderly (Prince et al 1999). It has also been used in low income and middle income countries in other regions (Castro-Costa et al., 2007; Prince et al., 2004). Its concurrent and criterion validity across 14 different settings in Europe was satisfactory (Prince et al., 1999). Most of the previous Euro-D studies suggests two common factors; affective suffering (including depression, tearfulness and wishing to death) and motivation (including loss of interest, poor concentration and lack of enjoyment). Evidence for its cross-cultural validity in developing countries including China, India, Latin America and Africa suggest a similar factor structure across these settings (Prince et al., 2004). There is as yet no evidence for the validity of Euro-D in Thailand.

In the current study we attempted to test Euro-D’s internal reliability, its concurrent validity against the diagnosis of major depressive episodes provided by the Mini International Neuropsychiatric Interview (MINI) (Sheehan et al., 1998) and its construct using principal component analysis on a sample of psychiatric clinic elderly attendees.

Method

Participants

We sampled consecutive patients attending a busy outpatient practice in a psychiatric hospital in the suburb of Eastern Bangkok. Patients aged 60 or above were approached in the waiting room when they attended the clinic to see their doctor. There were 8 psychiatrists participating in the study. A research worker, who was a postgraduate student in population and social research, approached the patients and asked for informed consent. Study was conducted according to the guidelines issued by the institutional review board of the hospital involved.

Procedures

After reading a study information sheet, those who agreed to take part were interviewed either before or after seeing their doctor in a private room on the premises. Those with evidence of severe cognitive impairment, dangerous behaviours, very frail, or with limitations of comprehension were excluded. We continued recruitment until our target of 150 completed interviews was met. The majority of the patients were presenting with common conditions such as anxiety, depression, dementia, and substance-related disorders.
Interviews with Euro-D were carried out before or after the participants saw their doctors. Each of the 8 psychiatrists administered MINI-Thai version, the major depressive episode section (Kittiratpaiboon & Wongkampin, 2004), while the trained research worker administered Euro-D. We blinded the MINI interviewers to the Euro-D responses obtained by the research worker and vice versa.

**Measures**

The Euro-D is a structured common depressive symptoms scale derived from the Geriatric Mental State-AGECAT (GMS-AGECAT) interview (Copeland, Dewey, & Griffiths-Jones, 1986), SHORT-CARE (Gurland, Golden, Teresi, & Challop, 1984) and other measures including the Centre for Epidemiological Studies Depression Scale (CES-D) (Radloff, 1977), Zung Self-Rating Depression Scale (ZSDS) (Zung, 1965), and the Comprehensive Psychological Rating Scale (CPRS) (Asberg & Schalling, 1979). It was designed to be administered by trained lay interviewers and only consists of 12 items dealing with: depression, pessimism, wishing death, guilt, sleep, interest, irritability, appetite, fatigue, concentration, enjoyment and tearfulness. A cut-point of 3/4 was identified using ROC analysis in studies carried out in 14 European countries and produced a range of sensitivity (63%-83%) and specificity (49%-95%). Principal components analysis generated two factors (affective suffering and motivation) that were common to nearly every participating European country (Prince et al., 1999) and for Indian, Latin-American and Caribbean centres (Prince et al., 2004). Internal consistency was universally satisfactory, ranging from 0.83-0.93 (Prince et al., 2004).

In Thailand, the Euro-D items appeared to cover symptoms recognised locally as common in psychological disorders in older adults. The original version was carefully translated and back-translated into English. First, a team of bilingual mental health professionals and social scientists developed the first translation, paying particular attention to conceptual and semantic equivalence. Two English speaking old age psychiatrist with training and extensive experience with the GMS assisted to answer any questions about the original version. The first translation was piloted on an urban community sample of 12 elderlies to test its clarity and comprehensibility. Some questions (e.g. sleep, irritability, appetite, weight loss) simply required translation. Others were adapted to ensure local equivalence or to aid comprehension (i.e. depression, concentration and pessimism). “Can you concentrate on entertainment or reading” was replaced with “Can you concentrate on daily activities that you like, such as listening to the monks’ teachings, the radio or watching television”. ‘Have you been feeling depressed?’ was replaced with ‘have you been feeling sad, gloomy or in despair’. It was most difficult to translate the question on pessimism. The original question asks ‘tell me your hopes for the future’ and rates pessimism if the person cannot describe at least one hope. In the Thai context we anticipated that older people, most of whom were Buddhist, would not mention any hopes because they were expected at their age to view life with contentment and to take each day as it comes. Asking about ‘future hopes’ might therefore elicit feelings of unfamiliarity and discomfort. This could lead to misinterpreting true pessimism. We changed the wording to ‘do you have any hope that in the near future good things might happen to you?’. We hoped that the modified closed question would enable to valid response.

The Mini International Neuropsychiatric Interview (MINI) was used as gold standard criterion measure. MINI is a short structured diagnostic interview, developed for DSM-IV and ICD-10 psychiatric disorders. The section on major depressive episode was used. The Thai version of MINI was validated against diagnoses made by local psychiatrists in a clinical setting (Kittiratpaiboon & Wongkampin, 2004) and used as a gold standard criterion against a newly developed screening test in a previous study (Arunpongpisal et al., 2006). It was also
used as a main diagnostic instrument in a recent national mental health survey (Department of Mental Health, 2003).

Data analysis
The output data from the MINI and Euro-D were read into STATA and the following parameters were assessed.
1) The prevalence of major depressive disorder according to each method of assessment.
2) The absolute numbers of participants in each cell of the cross tabulation by test and criterion status, that is true positive, false negative, false positive and true negative.
3) The kappa with 95% confidence intervals as the primary measure of agreement.
4) Coefficient of agreement (the proportion of participants classified as positive by either assessment that are classified as positive by both).
5) To assess concurrent validity, the scores of the Euro-D were compared with the diagnosis of major depressive disorder provided by the MINI. A positive case was defined as an individual who was diagnosed as having a major depressive episode. The area under the receiver operating curve (ROC), with the corresponding 95%CI, was estimated for major depressive disorder criterion. The optimal cut-off point was identified and sensitivity, specificity, positive, and negative predictive values were calculated.
6) The inter-item correlations and Cronbach’s alpha coefficients were calculated as measures of the internal consistency of the Euro-D.
7) A principal component analysis (PCA) of the Euro-D was performed, in an exploratory approach.

Results

Demographic and clinical characteristics
Over a two-month period we approached 167 patients of whom 150 completed the interview (response rate 89.8%). The demographic characteristics of the sample are shown in Table 1. The principal diagnoses were mood disorders (29.3%), anxiety disorders (23.3%), schizophrenia and delusional disorders (16.7%), organic mental disorders (16%), headache (4%), substance abuse and dependence (2.7%) and other disorders (7.3%).

Table 1. Sociodemographic characteristics (n=150)

<table>
<thead>
<tr>
<th>Variables</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Mean 69.4 (s.d.=7.0, range 59-88)</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>60</td>
<td>40</td>
</tr>
<tr>
<td>female</td>
<td>90</td>
<td>60</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>single</td>
<td>11</td>
<td>7.33</td>
</tr>
<tr>
<td>married</td>
<td>85</td>
<td>56.7</td>
</tr>
<tr>
<td>widowed</td>
<td>40</td>
<td>26.7</td>
</tr>
<tr>
<td>Divorced/separated</td>
<td>14</td>
<td>9.3</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 4 years</td>
<td>31</td>
<td>20.7</td>
</tr>
<tr>
<td>4 years (primary school)</td>
<td>81</td>
<td>54.0</td>
</tr>
<tr>
<td>&gt;4 years</td>
<td>38</td>
<td>25.3</td>
</tr>
</tbody>
</table>
**Scale distribution**
Table 2 shows the prevalence of each Euro-D symptom across the whole sample and the prevalence of each Euro-D symptom comparing patients who were diagnosed and not diagnosed with depression.
As shown in Table 2, the mean number of items rated as present during the past month per subject was 4.9 (95%CI 4.4-5.3). The frequencies of items ranged from 16.7% (concentration) to 58.7% (irritability). Only one item (pessimism) did not significantly discriminate patients depressed from those who were not. Over 75% of depressed patients reported that they had been depressed, had sleep trouble, been irritable, and felt fatigue.

**Table 2.** Euro-D item prevalences and scale scores (N=150)

<table>
<thead>
<tr>
<th>Symptom</th>
<th>N</th>
<th>Total %</th>
<th>Depressive patients diagnosed by MINI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Yes (n=51) %</td>
</tr>
<tr>
<td>Depression</td>
<td>76</td>
<td>50.7</td>
<td>79.7</td>
</tr>
<tr>
<td>Pessimism</td>
<td>85</td>
<td>56.7</td>
<td>64.4</td>
</tr>
<tr>
<td>Suicidality</td>
<td>40</td>
<td>26.7</td>
<td>50.9</td>
</tr>
<tr>
<td>Guilt</td>
<td>44</td>
<td>29.3</td>
<td>52.5</td>
</tr>
<tr>
<td>Sleep</td>
<td>84</td>
<td>56.0</td>
<td>76.3</td>
</tr>
<tr>
<td>Interest</td>
<td>47</td>
<td>31.3</td>
<td>54.2</td>
</tr>
<tr>
<td>Irritability</td>
<td>88</td>
<td>58.7</td>
<td>89.8</td>
</tr>
<tr>
<td>Appetite</td>
<td>64</td>
<td>42.7</td>
<td>71.2</td>
</tr>
<tr>
<td>Fatigue</td>
<td>86</td>
<td>57.3</td>
<td>93.2</td>
</tr>
<tr>
<td>Concentration</td>
<td>25</td>
<td>16.7</td>
<td>33.9</td>
</tr>
<tr>
<td>Enjoyment</td>
<td>40</td>
<td>26.7</td>
<td>44.1</td>
</tr>
<tr>
<td>Tearfulness</td>
<td>52</td>
<td>34.7</td>
<td>59.3</td>
</tr>
</tbody>
</table>

**Concurrent validity**
The concordance between Euro-D and MINI for major depressive episode was fair. The level of agreement for the sample (kappa) was 0.37 (95%CI 0.22-0.52). Euro-D tended to overdiagnosis with respect to MINI with a prevalence of 39.3% compared with 34% for major depressive disorder according to MINI. The area under ROC curve for the overall discriminability of the Euro-D scale against the criterion of MINI major depressive disorder was 0.78 (95%CI 0.70-0.85) (Figure 1). Inspection of psychometric indices at different cut-points suggests a cut-point of 5/6 with sensitivity of 84.3% and specificity of 58.6%. The positive and negative predictive values were 55.9 and 80.2, respectively (Table 3). A higher cut-point of 6/7 will yield lower sensitivity (64.7%) and higher specificity (73.7%) (Table 3).

**Table 3.** Performance of the Thai Euro-D in screening major depressive disorder in older patients

<table>
<thead>
<tr>
<th>Cut-off score</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Positive predictive value</th>
<th>Negative predictive value</th>
</tr>
</thead>
<tbody>
<tr>
<td>3/4</td>
<td>94.1</td>
<td>34.3</td>
<td>46.5</td>
<td>90.2</td>
</tr>
<tr>
<td>4/5</td>
<td>90.2</td>
<td>46.5</td>
<td>51.2</td>
<td>87.9</td>
</tr>
<tr>
<td>5/6</td>
<td>84.3</td>
<td>58.6</td>
<td>55.9</td>
<td>80.2</td>
</tr>
<tr>
<td>6/7</td>
<td>64.7</td>
<td>73.7</td>
<td>59.5</td>
<td>75.9</td>
</tr>
<tr>
<td>7/8</td>
<td>49.0</td>
<td>82.8</td>
<td>76.0</td>
<td>74.4</td>
</tr>
</tbody>
</table>
The Cronbach’s alpha for the total scale was sufficiently high at 0.72, although the internal consistency could be improved upon by the omission of EURO4 (pessimism), which marginally increased the standardised alpha to 0.74.

**Construct validity**

For factor analysis, the adequacy of the data for factor analysis was assessed by inspecting the correlation matrix, which showed that 10 coefficients in 66 were >0.3. Moreover, the Kaiser-Mayer-Olkin value was 0.77 and the Barlett’s test of sphericity achieved high level of statistical significance ($p<0.001$), supporting the factorability of the correlation matrix.

Principal component analysis revealed the presence of at least three components with eigenvalues exceeding 1, following inspection of the scree plot. Factor one conformed to affective suffering (depression, suicidality and tearfulness) with smaller contribution (0.4-0.5) from guilt, irritability and fatigue, and factor two indicated motivation (interest, concentration and enjoyment) but negligible loadings from other items. The affective suffering factor account for 26% of the variance in the items, and the motivation factors for 12%. The third factor, with Eigenvalue marginally over 1, was largely loaded on by sleep and appetite.
Table 4: Principal component analysis of Euro-D

<table>
<thead>
<tr>
<th></th>
<th>Factor1</th>
<th>Factor2</th>
<th>Factor3</th>
<th>Factor4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Eigenvalue</strong></td>
<td>3.1</td>
<td>1.5</td>
<td>1.1</td>
<td>1.0</td>
</tr>
<tr>
<td><strong>Variance (%)</strong></td>
<td>26</td>
<td>12</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Depression</td>
<td>0.65</td>
<td>0.03</td>
<td>0.34</td>
<td>-0.18</td>
</tr>
<tr>
<td>Pessimism</td>
<td>-0.03</td>
<td>0.08</td>
<td>-0.02</td>
<td>0.92</td>
</tr>
<tr>
<td>Suicidality</td>
<td>0.64</td>
<td>0.26</td>
<td>-0.02</td>
<td>0.04</td>
</tr>
<tr>
<td>Guilt</td>
<td>0.56</td>
<td>0.19</td>
<td>0.02</td>
<td>-0.22</td>
</tr>
<tr>
<td>Sleep</td>
<td>-0.03</td>
<td>0.06</td>
<td>0.79</td>
<td>-0.21</td>
</tr>
<tr>
<td>Interest</td>
<td>0.09</td>
<td>0.60</td>
<td>0.37</td>
<td>0.17</td>
</tr>
<tr>
<td>Irritability</td>
<td>0.42</td>
<td>0.23</td>
<td>0.43</td>
<td>-0.17</td>
</tr>
<tr>
<td>Appetite</td>
<td>0.24</td>
<td>0.01</td>
<td>0.69</td>
<td>0.28</td>
</tr>
<tr>
<td>Fatigue</td>
<td>0.55</td>
<td>0.20</td>
<td>0.34</td>
<td>0.17</td>
</tr>
<tr>
<td>Concentration</td>
<td>0.17</td>
<td>0.68</td>
<td>0.00</td>
<td>0.10</td>
</tr>
<tr>
<td>Enjoyment</td>
<td>-0.00</td>
<td>0.83</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Tearfulness</td>
<td>0.73</td>
<td>-0.12</td>
<td>-0.00</td>
<td>0.08</td>
</tr>
</tbody>
</table>

Discussion

Depression is an important condition affecting older people’s health and quality of life which needs a screening instrument to identify individuals at risk. Euro-D is a useful instrument which serves this purpose. To our knowledge, this is the first study to test the reliability and validity of Thai version of Euro-D. Attempts were made to ensure that content and semantic equivalence were achieved by strict translation and back-translation procedures with minor modifications to make the items more comprehensible and relevant to the Thai context. It was found to be feasible and easily administrable by trained interviewers in this clinical setting.

Our results show that the frequency of individual items rated as present during the previous month in this study was somewhat higher than those reported in a previous Euro-D studies in European countries (Prince et al., 1999). The prevalence of the pessimism item was particularly high and it did not discriminate well between the depressed and non-depressed groups. Although we altered the original version of the item it was still the case that few older people volunteered any hopes for the future. The Buddhist doctrines of contentment and and acceptance of the current life situation are likely to explain this view. The hope of better living in the future, for instance in terms of gaining material good, contradicts the traditional path toward merit in the afterlife (Pfanner & Ingersoll, 1962). Guilt was also highly prevalent. This may be due to people misunderstanding the guilt item or the tendency to volunteer feeling a burden on children and this being rated as a symptom. It is widely perceived among Thai older people that they should rely on their children when they get old and dependent (Knodel, Chayovan, & Siriboon, 1992), but many expressed guilt about this. Fatigue was also high, possibly because there were extra untreated physical illnesses or because many older Thai people are still working hard to make ends meet even after their retirement. Further studies using item response theory analysis to explore for differential item functioning (DIF) may be helpful to investigate whether there is item bias in the Thai version and to explain the high levels of these symptoms. The area under the ROC curve for the overall discriminability of the Euro-D scale against the criterion of any MINI depressive disorders was 0.78 (0.70-0.85). This rather impressive degree of overall prediction, coupled with the high sensitivity (94%) and poor specificity (34%) of the Euro-D cut-point of 3/4 suggested that this cut-point is too low for the Thai clinical sample. A cut-off score of 5/6 had lower sensitivity (84.3%)
but higher specificity (58.6%) to screen major depressive episode compared with the usual cut
off reported previously. This higher cut-off might reflect more depressive symptoms among
psychiatric patients. The majority of cases of depressive disorders in clinical settings
obviously tend to be more severe compared with cases in general population. In community
settings, cases may tend to have lower screening score. Further validation studies in the
community setting will be needed. Taking account of the Thai context and improving the
clarity of the wordings may help to improve the overall sensitivity and specificity.

The Cronbach alpha coefficient shows a high internal consistency for the whole Euro-
D questionnaire, although the item on pessimism would appear to improve the overall internal
consistency if deleted. As discussed earlier, the concept of pessimism may be different among
Thai elderly. Finding a better way to address the issue of pessimism may help to improve the
overall psychometric properties.

One of the questions raised in a validation study is about what is the real construct
behind the instrument to be validated. The Euro-D was mainly derived from GMS/AGECAT
and designed to be used by trained lay interviewers, whereas MINI was based on ICD-10 and
DSM-IV criteria and used by mental health professionals as a diagnostic tool for major
depressive episode These discrepancies may provide an explanation for the fair kappa for the
agreement between Euro-D and MINI cases.

It is striking that the internal structure of the Euro-D as suggested by principal
components analysis is consistent with most of the previous studies (Castro-Costa et al., 2007;
Prince et al., 2004). Our version of the Euro-D also shares two common constructs, the first
one being affective suffering, including depression, tearfulness and wishing to die, and second
one being motivation including loss of interest, poor concentration and lack of enjoyment.
They were mapped well onto the two common main factors reported previously, although
there were slight differences in the individual items included in each factor (Castro-Costa et
al., 2007; Prince et al., 2004; Prince et al., 1999). The two extra factors were identified in our
study, one including sleep and appetite and the other one including pessimism. The three
items had eigenvalues of about 1. This may not be so surprising as three- or four-factor
structure was also reported in some of the European settings. For example, sleep and appetite
were found to be loading on another factor in Sweden, Netherlands, and Iceland and
pessimism loading on a separate factor in Germany (Prince et al., 1999). Although the four-
factor structure is somewhat different, it remains difficult to establish that it does vary from
the European studies. Confirmatory factor analysis may be a useful technique to test whether
the four–factor model has better overall goodness-of-fit than the two-factor solution.

There were some limitations in this study. First, the study was conducted in a
psychiatric hospital on a sample of patients with a variety of mental conditions. The study
sample was obviously different from those based in communities or general practice. Second,
the sample size available for this validation study was relatively small. More studies with a
larger sample size in a variety of clinical settings are warranted. In addition, item bias cannot
be excluded as an explanation for the relatively high prevalences of many Euro-D items in our
study. Further studies in cross-cultural settings using item response theory analysis should
provide more clues.

In conclusion, this is the first study to test the reliability and validity of Thai version of
Euro-D. This validation study reports the use of Euro-D in a clinical setting. The instrument
has demonstrated rather different properties to those shown in previous studies. The different
cut-point found in the current study suggests the need for caution in interpretation of Euro-D
scores in Thailand. Further studies are still needed to validate Euro-D in the Thai community.
Acknowledgements

This research was supported by the Wellcome Trust, UK (WT 078567). We are grateful to the patients who participated and psychiatrists and staff at Kalayarajanakarin Institute for their assistance in facilitating the study. We are particularly indebted to Drs. Kampanart Tansitabudkul and Wanatda Tomkapanit. We thank Professor Martin Prince for his help with the Euro-D.

References


