Performance of the CES-D and its Short Forms in Screening Suicidality and Hopelessness in the Community

YIN BUN CHEUNG, RSW, PHD, KA YUET LIU, M.PHI, and PAUL S.F. YIP, PHD

A measurement scale should be short and quick to complete if it is to be practically useful. Drawing on data from a community-based survey of 2,178 people in Hong Kong, we compared five short forms (5- to 10-item) and the original version (20-item) of the Center for Epidemiologic Studies–Depression Scale (CES-D; Radloff, 1977) in predicting suicidal attempts and suicidal thoughts. Short forms with as few as nine items performed in ways very similar to the full version; a version with only five items had a detectable difference from the full version. Sensitivity, specificity, and predictive values in differentiating people with and without suicidal thought or attempt change almost linearly with the cut-offs.

Recent studies of global disease burden indicate that depression, left untreated, is the single most important source of burden in terms of days lost to illness (Murray & Lopez, 1996; World Health Organization [WHO], 2001). The lifetime suicide risk among depressed patients can be as high as 6% (Inskip, Harris, & Barracough, 1998), though the risk level depends on treatment setting (Bostwick & Pankratz, 2000). Depression, however, is often undetected and therefore untreated in the community and primary care setting (Boey, 1999; Sliman, Donohue, Jarjoura, & Ognibene, 1992), which makes effective screening for depression an important public health issue. The Center for Epidemiologic Studies–Depression Scale (CES-D), developed by Radloff (1977), is a popular screening tool used in many countries and has gone through numerous validation exercises. Given the high suicide mortality among depressed patients, information about the characteristics of the CES-D in relation to suicidal thoughts and behavior is of great importance. Moreover, useful though the CES-D is, the length of 20 items is a factor that hinders more widespread use. There is a concern that the CES-D can be stressful for some respondents, which leads to a suggestion to reduce this stress by using shorter versions of the instrument (Kohout, Berkman, & Evans, 1993). The present study aims to shed light on the characteristics of the full version and several proposed short forms of the CES-D in relation to screening suicidal attempt and suicidal thought. A secondary aim is to examine the choice of different cut-points to identify people as suicidal.

METHODS

Survey Design

This is the first territory-wide survey on nonfatal suicidal behavior using commu-
nity-based samples in Hong Kong. The survey was conducted during the first half of 2004 among the local resident population (between 15 and 59 years of age) of Hong Kong. All domestic helpers from overseas countries with conditional working visas are excluded from the targeted population. The sampling frame employed was based on the Frame of Quarters maintained by the Census and Statistics Department, which facilitates the most complete and up-to-date register of residential addresses in Hong Kong. From each household contacted, we attempted to use the last birthday method to recruit one respondent. Informed written consent was obtained from the participant before the interview. The survey consisted of two parts. Firstly, trained interviewers conducted a face-to-face interview, which included items on sociodemographic background, the CES-D, and the Beck's Hopelessness Scale (BHS; Beck et al., 1974; Shek, 1993), among other things. Upon completion of the interview, the respondents were invited to self-complete a questionnaire on more sensitive topics, including suicidal ideation and suicide attempt. A total of 3,609 households were contacted; 1,389 households did not participate either because they refused to participate in the selection of individuals or the selected individuals refused to participate. A total of 2,219 persons aged between 15 and 59 participated in the study. The response rate was 61.5% (2,219/3,609). Given a sensitive topic like suicidal behavior, this response rate was satisfactory. Moreover, the demographic profile of this sample was found to be similar to those of the Hong Kong general population of that age range. A very small number \((n = 41)\) of respondents were excluded from this analysis because of missing values in the variables concerned. Therefore, the final sample size is 2,178.

**Measures**

**CES-D.** The CES-D contains 20 items. The answers to each item, which indicate the frequency of the symptoms, are scored as ranging from 0 (rarely or none of the time) to 3 (most or all of the time). The total scores can range from 0 to 60. Radloff (1977) suggested and it is commonly accepted that persons who score 16 or above are likely to be clinically depressed; however, some studies suggest that the cut-point needs to be raised to 24 or higher if the tool is to achieve appropriate diagnostic characteristics (Gotlib, Lewinsohn, & Seeley, 1995; Roberts, Lewinsohn, & Seeley, 1991). A Chinese version of the CES-D has been validated and used in Hong Kong (Cheung & Bagley, 1998; Lam et al., 2004).

Several short forms of the CES-D have been developed. Cole, Rabin, Smith, and Kaufman (2004) commented that the 9-item short form developed by Santor and Coyne (1997) and the 10-item and 11-item short forms developed by Kohout et al. (1993) were methodologically more rigorous than others. Santor and Coyne selected the items according to their ability to discriminate between groups. Kohout et al. selected items based on factor analytic results, but did not indicate superiority on the part of the 11-item version as compared to the 10-item version. Cole et al. developed a new 10-item version based on the Rasch model. They suggested that this instrument may perform better than the previous ones in terms of detecting a range of severity of depression. To our knowledge the shortest form of the CES-D is the 5-item version developed by Shrout and Yager (1989) using discriminant analysis. Andresen, Malmgren, Carter, and Patrick (1994) developed another 10-item version of the CES-D by analysis of item-total correlation. This approach to the shortening of measurement scale is generally not recommended because of a problem of correlated errors (Cheung et al., 2003; Coste, Guillemot, Pouchoit, & Fermanian, 1997), such that items with large amount of error may be selected as a result. Nevertheless, this short form has been used in Hong Kong in two studies of the elderly (Boey, 1999; Cheng & Chan, 2005).

In addition to the full version of the CES-D, we used the 10-item version of Cole et al. (2004; CES-D-10-Cole), the 10-item
version of Andresen et al. (1994; CES-D-10-Andresen), the 10-item version of Kohout et al. (1993; CES-D-10-Kohout), the 9-item version of Santor and Coyne (1997; CES-D-9), and the 5-item version of Shrout and Yager (1989; CES-D-5). The scores on the five short forms of the CES-D were calculated based on the 20 items of the CES-D.

Suicidal Thoughts and Attempts. The self-administered questionnaire included a question “During the past 12 months, had you ever attempted to commit suicide?” and a question “During the past 12 months, had you ever considered suicide?” Respondents who gave positive replies to the first and second question were considered to have suicidal attempt and suicidal thought, respectively.

Beck’s Hopelessness Scale. The Chinese version of the BHS was translated and validated in Hong Kong (Shek, 1993). A slight modification in the design of the Chinese scale is that instead of asking the respondents to give a Yes/No answer, which was considered narrow in the response range, the respondents were asked to respond Strongly Agree, Moderately Agree, Slightly Agree, Slightly Disagree, Moderately Disagree, or Strongly Disagree to the items. The answers are coded as 1 to 6 in a way that a higher value represents a higher level of hopelessness. Scores on the Chinese BHS therefore range from 20 to 120.

Statistical Analysis

We performed receiver operating characteristics (ROC) analysis using the scores on the CES-D and the short forms to detect self-report of suicide attempts in the last 12 months and suicidal thoughts in the last 12 months. A lack of predictive power is indicated if the 95% confidence interval (CI) of the area under ROC curve (AUC) included the null value of 0.5; an AUC closer to 1 indicated a better predictive power (Hanley & McNeil, 1982). A nonparametric procedure was used to test the equality of the AUCs given by the correlated instruments (DeLong, DeLong, & Clarke-Pearson, 1988). Pearson’s correlation coefficients and their 95% confidence intervals were calculated for the association between versions of CES-D and the BHS. The risk of suicidal attempts and suicidal thoughts was assessed in relation to the CES-D scores by the locally weighted regression smooth method (Cleveland, 1979). The relation between the CES-D and the short forms were assessed by ordinary least square regression, and the linearity examined graphically. The positive and negative predictive values (PPVs and NPVs) of the previously recommended cut-points of 16 and 24 as well as a wide range of other cut-points on the CES-D were also assessed. The statistical package STATA (Version 8.0; College Station, Texas: StataCorp) was used to perform the analysis.

RESULTS

Out of 2,178 respondents, 38 reported suicidal attempts and 147 suicidal thoughts in the last 12 months. The prevalence rates (95% CI) were 1.7% (1.2 to 2.4%) and 6.7% (5.7 to 7.9%), respectively. The mean CES-D score was 8.1; 7.5% of the respondents had CES-D score ≥ 24. Other descriptive statistics are also given in Table 1. Figure 1 shows the smoothed probability of suicidal attempts (solid line) and suicidal thoughts (dashed line) in the last 12 months in relation to CES-D score. The cut-points of 16 and 24 are superimposed on the figure. Suicidality increased quite steadily and roughly linearly in the region of CES-D score larger than 10.

Table 2 tabulates the AUCs of the

<table>
<thead>
<tr>
<th>Statistics</th>
<th>CES-D scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>0</td>
</tr>
<tr>
<td>25th percentile</td>
<td>1</td>
</tr>
<tr>
<td>Median</td>
<td>5</td>
</tr>
<tr>
<td>75th percentile</td>
<td>12</td>
</tr>
<tr>
<td>Maximum</td>
<td>60</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>8.1 (9.6)</td>
</tr>
<tr>
<td>% ≥ 16</td>
<td>16.6</td>
</tr>
<tr>
<td>% ≥ 24</td>
<td>7.5</td>
</tr>
</tbody>
</table>

TABLE 1
Descriptive Summary of the Full Version
CES-D Scores (N = 2,178)
CES-D and those of the five short versions in detecting suicidal attempts in the last 12 months. The CES-D achieved an AUC (95% CI) of 0.810 (0.735 to 0.885). The CES-D-5 gave the largest AUC, at 0.824 (0.751 to 0.896). Nevertheless, the null hypothesis of no difference in the AUCs between the instruments was rejected ($p < 0.001$). Pairwise comparison showed that the CES-D had AUC significantly larger than that of the CES-D-10-Cole ($p = 0.003$) and that of the CES-D-10-Andresen ($p = 0.048$). There was no significant difference in the other pairwise comparisons (each $p > 0.05$).

All versions of the CES-D were associated with the BHS to a similar degree. The correlation coefficients were in the neighbourhood of 0.5 (Table 3). Apart from the

<table>
<thead>
<tr>
<th>Versions of CES-D</th>
<th>AUC</th>
<th>95% CI</th>
<th>AUC</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>CES-D (20-item)</td>
<td>0.810</td>
<td>0.735 to 0.885</td>
<td>0.814</td>
<td>0.775 to 0.852</td>
</tr>
<tr>
<td>CES-D-10-Cole</td>
<td>0.772</td>
<td>0.690 to 0.854</td>
<td>0.791</td>
<td>0.749 to 0.832</td>
</tr>
<tr>
<td>CES-D-10-Kohout</td>
<td>0.811</td>
<td>0.734 to 0.887</td>
<td>0.810</td>
<td>0.771 to 0.850</td>
</tr>
<tr>
<td>CES-D-10-Andresen</td>
<td>0.798</td>
<td>0.722 to 0.873</td>
<td>0.803</td>
<td>0.764 to 0.842</td>
</tr>
<tr>
<td>CES-D-9</td>
<td>0.806</td>
<td>0.732 to 0.879</td>
<td>0.805</td>
<td>0.765 to 0.844</td>
</tr>
<tr>
<td>CES-D-5</td>
<td>0.824</td>
<td>0.751 to 0.896</td>
<td>0.805</td>
<td>0.754 to 0.837</td>
</tr>
</tbody>
</table>

TABLE 2
Area Under Receiver Operating Characteristics Curve for Detecting Suicidality in the Last 12 Months

Figure 1. Risk of suicidal attempts (solid line) and suicidal thoughts (dashed line) in relation to CES-D score obtained by locally weighted regression smooth. Vertical reference lines indicate CES-D scores of 16 and 24.
Moving the cut-points upward to 24 substantially reduced the sensitivity to Pearson’s Correlation Coefficient ($r$) Between Version of CES-D and the Beck’s Hopelessness Scale, although the specificity and the PPV are improved. Using the short versions of the CES-D with the aforementioned cut-points that correspond to CES-D 16 and 24 gave very similar results; using other cut-points obtained from the aforementioned regression equations also gave results very similar to those shown in Figure 3 (details not shown).

The findings on sensitivity, specificity, and NPVs using the CES-D scores of 16 and 24 to identify people with suicidal thoughts in the last 12 months were not very different from those concerning the detection of suicidal attempts (Figure 4). Nevertheless, the PPVs were much larger. When the cut-point of 16 was used, the PPV was 25.5%. When the higher cut-point of 24 was used, the PPV increased substantially to 36.2%. Again, using the short versions with the aforementioned cut-points or other cut-points based on the regression equations gave similar results (details not shown).

Taking into account the aforementioned findings on suicidality and hopelessness, we decided to exclude the 10-item versions of Cole et al. (2004) and Andresen et al. (1994) and the 5-item version from further analysis. Figure 2 gives scatter plots of the scores on the CES-D versus that of the two remaining short versions. Linear regression lines are superimposed on it, and the regression equations are given on the top of each panel. The relations between scores on the CES-D and the short versions are very linear. Using the regression equations, we calculated that a score of 9 and a score of 14, respectively, on the CES-D-10-Kohout correspond to 16 and 24 on the CES-D. For instance, the regression equation $y = 0.316 + 1.750 \times$ (given on top of the upper panel in Figure 2) indicates that a CES-D-10-Kohout score that predicts a CES-D score of 16 was obtainable by $(16 - 0.316)/1.750 = 9$. Similarly, a score of 8 and 13 on the CES-D-9 corresponds to 16 and 24 on the CES-D. The $R^2$ of the linear regression models were 93.3% and 92.3% for the CES-D-10-Kohout and CES-D-9, respectively.

Figure 3 shows the characteristics of using a wide range of CES-D cut-points for detecting suicidal attempt and suicidal thought. Using the conventional cut-point of 16, the CES-D has a moderate level of sensitivity (65.8%) in identifying people with suicidal attempts. Moving the cut-points upward to 24 substantially reduced the sensitivity to 36.8%, although the specificity and the PPV are improved.Using the short versions of the CES-D with the aforementioned cut-points that correspond to CES-D 16 and 24 gave very similar results; using other cut-points obtained from the aforementioned regression equations also gave results very similar to those shown in Figure 3 (details not shown).

Taking into account the aforementioned findings on suicidality and hopelessness, we decided to exclude the 10-item versions of Cole et al. (2004) and Andresen et al. (1994) and the 5-item version from further analysis. Figure 2 gives scatter plots of the scores on the CES-D versus that of the two remaining short versions. Linear regression lines are superimposed on it, and the regression equations are given on the top of each panel. The relations between scores on the CES-D and the short versions are very linear. Using the regression equations, we calculated that a score of 9 and a score of 14, respectively, on the CES-D-10-Kohout correspond to 16 and 24 on the CES-D. For instance, the regression equation $y = 0.316 + 1.750 \times$ (given on top of the upper panel in Figure 2) indicates that a CES-D-10-Kohout score that predicts a CES-D score of 16 was obtainable by $(16 - 0.316)/1.750 = 9$. Similarly, a score of 8 and 13 on the CES-D-9 corresponds to 16 and 24 on the CES-D. The $R^2$ of the linear regression models were 93.3% and 92.3% for the CES-D-10-Kohout and CES-D-9, respectively.

Figure 3 shows the characteristics of using a wide range of CES-D cut-points for detecting suicidal attempt and suicidal thought. Using the conventional cut-point of 16, the CES-D has a moderate level of sensitivity (65.8%) in identifying people with suicidal attempts. Moving the cut-points upward to 24 substantially reduced the sensitivity to 36.8%, although the specificity and the PPV are improved. Using the short versions of the CES-D with the aforementioned cut-points that correspond to CES-D 16 and 24 gave very similar results; using other cut-points obtained from the aforementioned regression equations also gave results very similar to those shown in Figure 3 (details not shown).

The findings on sensitivity, specificity, and NPVs using the CES-D scores of 16 and 24 to identify people with suicidal thoughts in the last 12 months were not very different from those concerning the detection of suicidal attempts (Figure 4). Nevertheless, the PPVs were much larger. When the cut-point of 16 was used, the PPV was 25.5%. When the higher cut-point of 24 was used, the PPV increased substantially to 36.2%. Again, using the short versions with the aforementioned cut-points or other cut-points based on the regression equations gave similar results (details not shown).

Santor and Coyne (1997) suggested to dichotomize the item responses by coding the two lower categories of reply as 0 and the higher two categories as 1. In a supplementary analysis we used this dichotomization for the CES-D-9-Santor and the results are practically the same as those reported above without the dichotomization (details not shown).

**DISCUSSION**

The screening of depression in health care, educational, and other settings has been widely advocated. Particularly, many national strategies and guidelines issued by professional bodies have included early detection of depression as a key measure to prevent suicide (e.g., American Psychiatric Association, 2003; U.K. Department of Health, 2001; U.S. Department of Health and Human Services, 2001). The CES-D is a popular instrument for the screening of depressive symptoms. Although many validation studies have been conducted on the CES-D, this study is...
Figure 2. Scatter plots of scores on the original CES-D versus scores on short versions of the CES-D. Linear regression lines are added, with the equation given on top of each panel. Top panel: CES-D-10 of Kohout et al. (1993); bottom panel: CES-D-9 of Santor and Coyne (1997). Jittering was applied to improve visual clarity.

the first to assess its diagnostic characteristics in detecting suicidal thoughts and behaviors. Moreover, compared to clinical or college student samples, our large community sample can yield information that is directly relevant to the screening and detection of depression among the general population. The CES-D is an epidemiological, not clinical, tool. We interpret and discuss the findings from this point of view. One should be cau-
Figure 3. Sensitivity (dotted line), specificity (long dashed line), positive predictive values (solid line), and negative predictive values (short dashed line) for detecting suicidal attempt by CES-D cut-points.

tious in extrapolating the present findings and recommendations to the clinical assessment area.

Questionnaires should be short and quick to complete if they are to be clinically useful (Higginson & Carr, 2001). This is especially important if the questions can be emotionally stressful, as is the case of the screening for depression (Kohout et al., 1993). Assessment of health and psychological constructs often requires composite measurement scales with high level of internal

Figure 4. Sensitivity (dotted line), specificity (long dashed line), positive predictive values (solid line), and negative predictive values (short dashed line) for detecting suicidal thought by CES-D cut-points.
consistency. This implies that the items are correlated and therefore the information they provide overlaps. Previous methodological studies suggested that the number of items in some scales can be reduced by about 70% without substantially compromising the measurement properties (Anne-Moran, Guyatt, & Norman, 2001; Shrout & Yager, 1989; Ware, Kosinski, & Keller, 1996).

In this study we examined five short forms of the CES-D. The major finding is that they performed in ways similar to the CES-D. The AUCs of all the short versions were similar to that of the CES-D in detecting suicidal attempts, suicidal thoughts, and BHS scores. Those of the CES-D-10-Cole, the CES-D-10-Andresen, and the CES-D-5 were only slightly inferior in relation to detecting suicidality. Cole et al. (2004) developed the 10-item version with the purpose of assessing a wide range of severity of depression. This is not the focus of ours. That of Andresen et al. (1994) is methodologically less ideal as it was derived from item-total correlation. The 5-item version was a bit weaker in association with BHS. Nevertheless, the main reason we dropped these three instruments from further analysis was to keep this article short and simple. In terms of detecting people with suicidal risk, the use of the 10-item version of Kohout et al. (1993) and the 9-item by Santor and Coyne (1997) did not appear to be disadvantaged in comparison to the CES-D.

These short forms are highly and linearly associated with the CES-D, the $R^2$ values were about 0.92 or 0.93. We can map the CES-D scores onto the short form scores by using the regression equations provided in this manuscript. Their sensitivity, specificity, PPV, and NPV in screening suicidal attempts and thoughts were practically the same.

The CES-D cut-point of 16 has been criticized by other researchers as being too loose and as wrongly identifying too many nondepressed people as likely to be depressed. Alternative cut-off values of 24 or higher have been recommended (e.g., Gotlib et al., 1995; Roberts et al., 1991). Such an increase in the cut-off value can improve the PPV; however, we have seen that the risk of suicidal attempts and thoughts increase quite steadily and roughly linearly in the region of CES-D scores larger than 10. This implies that any increase in the cut-off to improved PPV and specificity will unavoidably compromise the sensitivity and NPV. We do not see an “optimal” cut-off. The search for such a value may turn to be illusive; instead, one has to choose a cut-off that fits one’s purpose. Ideally the choice should be guided by not only a balance between diagnostic characteristics, but also the cost and benefit of, for example, successfully identifying a case and failure to identify a case. Unfortunately, such cost and benefit information is rarely available. So we will need to be careful in making any changes to the conventional cut-point. The information provided in this article should help to make informed choices according to the specific needs of each clinical or research team.

A limitation of the present study is that we did not actually administer the various short forms of the CES-D. Instead, the full version of the CES-D was administered and the responses to the 20 items were used to calculate the scores for the short form. Recent studies of a short version of a quality of life questionnaire found that the short questionnaire administered alone gave results very comparable to that given by the same items administered as part of the full version of the questionnaire (Cheung et al., 2004; Cheung, Khoo, Thumboo, Wee, & Goh, 2005). This suggests that the context of responding to the totality of an instrument does not always influence the responses to individual items. Nevertheless, we recommend further studies using the short forms of CES-D to verify if this is true for this particular measure. Other limitations are that suicidality was based on self-report and that the study population is limited to ethnic Chinese. Further studies using other criteria to assess the performance of the long and short forms of CES-D in this and in other populations, possibly using a randomized design to assign short and long forms to participants, will be useful.
REFERENCES


Santor, D. A., & Coyne, J. C. (1997). Shortening the CES-D to improve its ability to...


Manuscript Received: January 30, 2006
Revision Accepted: June 22, 2006