Validity of the Center for Epidemiologic Studies Depression Scale as a Screening Instrument of Major Depressive Disorder Among Japanese Workers

Koji Wada, MD, MSc1,3; Katsutoshi Tanaka, MD, PhD2; Gilles Theriault, MD, DrPH,3 Toshihiko Satoh, MD, MPH, PhD4; Masaru Mimura, MD, PhD5; Hitoshi Miyaoka, MD, PhD6; and Yoshiharu Aizawa, MD, PhD4

Background The Center for Epidemiologic Studies Depression Scale (CES-D) is used at workplaces to screen depressive disorders. The aim of this study was to examine the validity of the CES-D for depression in a workplace.

Methods The CES-D was administered to 2,219 workers (84.2% men; age 21–68 years) at a manufacturing company in Japan. Concomitantly all workers had an interview with the Mini International Neuropsychiatric Interview (MINI) as a gold standard for diagnosing major depressive disorder (MDD). The validity was evaluated by a receiver operating characteristic (ROC) curve.

Results The area under the ROC curve of the CES-D was 0.96 [95% Confidence Interval (CI): 0.94–0.99]. The optimal cut-off score of MDD was 19 for screening.


KEY WORDS: depression; screening; worker; CES-D; Japan

INTRODUCTION

Depression is a common health problem in workplaces. Depression is expected to be the second leading cause of disability worldwide in 2020 [Murray and Lopez, 1997]. The cost related to depression would be high. Greenberg et al. [2003] assessed the economic impact of depression at approximately $51.5 billion in the United States in 2000 due to absenteeism from work and lost productivity.

Companies have taken a variety of measures to minimize this problem such as screening, organizational intervention, and stress management program [Michie, 2002]. Screening for depression could identify some emotionally distressed employees and could improve outcomes in workplaces, particularly when screening is followed by adequate treatment and follow-up [U.S. Preventive Services Task Force, 2002; American College of Occupational and Environmental Medicine, 2003].

Self-reported questionnaires are inexpensive screening tools compared to interviews. The questionnaire known as the Center for Epidemiologic Studies Depression Scale (CES-D) is widely used for identifying people with depressive symptomatology, because its literacy is simple.
and it takes only 3–5 min to fill out [Radloff, 1977]. Even though it was developed to measure and detect depressive symptoms in the community population, it is also used as a screening tool for different populations such as first-visit patients to psychiatric clinics, adolescents, elders, and workers [Schulberg et al., 1985; Iwata et al., 1989; Furukawa et al., 1997; Haringsma et al., 2004; Yang et al., 2004].

The validity of CES-D has been tested in different communities and appears to be sound enough to be applied to patients, adolescents, and elders [Roberts and Vernon, 1983; Furukawa et al., 1997; Haringsma et al., 2004; Yang et al., 2004]. However, the validity of CES-D has not yet been studied in occupational health settings. While the receiver operating characteristic (ROC) analysis has been used to show the validity of CES-D in first-visit psychiatry patients and in the elders [Furukawa et al., 1997; Haringsma et al., 2004; Yang et al., 2004], it has not been used in workplaces.

The objectives of this study were to determine the validity of the self-reported CES-D in a workplace and to set the optimal cut-off score of the CES-D among Japanese workers.

**MATERIALS AND METHODS**

**Participants**

The participants of this study were workers who undertook the periodical examination between January and March 2005 in a Japanese company. In Japan, it is mandatory for employees to undertake a periodical examination once a year. Of 2,480 eligible workers, 62 did not take the health examination, 190 answered the questionnaire on depression but refused to participate in this study, 9 were temporarily suspended from their job (5 for mental disorders and 4 for other diseases). Thus 2,219 workers (1,868 men and 351 women) were enrolled in the study.

**Procedures**

A self-report questionnaire of CES-D was sent to all workers who took the periodical health examination (2,409 workers). Concomitantly, a Japanese version of the section of the Mini International Neuropsychiatric Interview (MINI) that addressed major depressive episode was administered to all workers [Otsubo et al., 2005]. Twelve public health nurses who had been trained by a psychiatrist participated in data collection over a 2-week period. Interviewers were blinded to the responses given to the CES-D.

**Measures**

The MINI is a short structured diagnostic interview of the Structured Clinical Interview for DSM-IV (SCID) [Lecrubier et al., 1997; Sheehan et al., 1997; Amorim et al., 1998; Sheehan et al., 1998]. The MINI includes 17 axis I categories in a format and has good correlation with SCID. In this study, we used the part of the MINI that identifies major depressive episode. Major depressive disorder (MDD) is defined as a score of 5–9 on that part of major depressive episode of the MINI.

The CES-D consists of 20 items with a four-point rating scale ranging from “0” (never or few) to “3” (usually). Symptoms score is the sum of the 20 items, in which the scoring was reversed for the items 4th, 8th, 12th, and 16th. The CES-D scores range from 0 to 60, with higher scores indicating high severity of depressive symptoms. The Japanese version of CES-D was translated from English and was found to be applicable to Japanese adults in the community [Shima et al., 1985].

**Statistics**

The sensitivity, specificity, positive predictive values, and negative predictive values were calculated for different cut-off scores of the CES-D to construct a ROC curve. Using the ROC analysis [Swets, 1988], the association between CES-D scores and the diagnosis of MDD as defined by the MINI was calculated. The ROC analysis is a method that evaluates the efficacy of a diagnostic test with continuous variables. The area under the curve (AUC) can be used to address the performance of a test. An AUC of 1.0 indicates perfect accuracy, while an AUC of 0.5 indicates a non-discriminating test. The software used for ROC analysis was the Statistical Package for Social Sciences (SPSS) 10.0J [SPSS, 1999].

The ROC curve is also used to determine the score that maximizes the efficacy of a screening test. The method used to define the optimal cut-off score is based on the distance between the ROC curve and the intersection of the sensitivity and 1-specificity on the ROC graph. A perfectly valid test has a cut-off score that passes by the intersection. When sensitivity and 1-specificity is away from 1.0, the ideal cut-off score is the point of the ROC curve located at the shortest distance to the intersection [Warner, 2004].

**Ethics**

The study protocol was reviewed and approved by the Committee of Human Research at Kitasato University, Japan. Written informed consent was obtained from each participant and confidentiality was protected.

**RESULTS**

The participants comprised 1,868 men and 351 women for a total population of 2,219 and this represented 89.5% of the eligible population. The demographic information
is presented in Table I. The mean age of participants was 42.0 years (SD = 11.4), range 21–68. The frequency of CES-D distribution by gender is shown in Figure 1. The percentage of participants with a CES-D score over 19 was 9.5%.

In order to construct an ROC curve, sensitivity and specificity were calculated for several scores of the CES-D. The sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) are shown in Table II. The sensitivity varied from 95.1% to 85.3% and the specificity varied from 82.2% to 93.1% in the central range of the curve. With a cut-off score of 16 which is the traditional dichotomized score in the literature, the sensitivity was 95.1% and the specificity was 85.0%. This yielded an elevated rate of false positives 89.3%. The sensitivity and the NPV were similar for cut-off scores from 15 to 17. At a cut-off score of 20, there was an obvious loss of sensitivity.

The ROC curve for CES-D is shown in Figure 2. The cut-off value determined as the shortest distance between any point in the ROC curve and the upper left intersection of the sensitivity and 1-specificity of the ROC graph was 19. The AUC calculated with ROC analyses was 0.96 [SE = 0.01; 95% Confidence Intervals (CI) 0.94–0.99; P < 0.001].

**DISCUSSION**

The objective of this research was to test the validity of the CES-D for MDD diagnosed by the MINI in a cohort of workers. The results suggest that the CES-D has a satisfactory validity for MDD. Concerning the overall validity of a test, an AUC of 0.5–0.7 indicates low accuracy, 0.7–0.9 moderate accuracy, and over 0.9 high accuracy [Swets, 1988]. According to this scale, the CES-D with an AUC value of 0.96 has a high validity in detecting MDD among workers. Furukawa et al. [1997] reported a moderate validity of CES-D applied to first-visit psychiatric patients and Yang et al. [2004] showed a high validity of CES-D in a group of adolescents in Taiwan.

The specificity and sensitivity depend on the cut-off scores used to define a disease. A more sensitive test tends to
be less specific and a more specific test tends to be less sensitive. Several screening instruments for depression are available and most of them have relatively good sensitivity but only fair specificity [U.S. Preventive Services Task Force, 2002]. Our results are in line with this.

The PPV values depend on the prevalence of disease. In this study, the prevalence of MDD is 1.8%. It is a relatively low prevalence, therefore the corresponding PPV would have a low value.

As shown in our analysis, the optimal cut-off score in our worker study is 19. The cut-off score of 19 (Sensitivity = 92.7%, Specificity = 91.8%, PPV = 17.6%, NPV = 99.9%) exhibited the highest sum of sensitivity and specificity. At a cut-off score of 19, the number of false positives is reduced with a minimum loss of sensitivity. Above this cut-off value, sensitivity decreased significantly without any further gain in specificity. The cut-off of 17 (Sensitivity = 95.1%, Specificity = 88.0%, PPV = 13.0%, NPV = 99.9%) had higher sensitivity, however, the number of false positives was higher. A cut-off score of 17 could be chosen if one is ready to accept more false positives.

The optimal cut-off score of CES-D varies in different populations. The optimal cut-off score for first-visit psychiatry patients in Japan whose prevalence rate of depression was 38.9% was 26 if greater emphasis is placed on sensitivity, 31 if equal weight is placed on sensitivity and specificity, and 34 if specificity is deemed more important [Furukawa et al., 1997]. The optimal cut-off score for older Chinese was 22 (Sensitivity = 0.75, Specificity = 0.51) using the diagnosis made by a physician as the gold standard [Cheng and Chan, 2005]. This cut-off score has a low sensitivity and a low NPV.

The traditional CES-D cut-off score which Radloff proposed in the United States was 16 (Sensitivity = 0.95, Specificity = 0.29) for the general population [Radloff, 1997]. In Japan, the optimal cut-off score was also proposed as 16 [Shima et al., 1985]. There might be two reasons why the optimal cut-off score of this study is higher than in these previous studies. First, the validity of CES-D was evaluated by the interview using the MINI. Second, the prevalence of MDD among workers could be low since workers are a selection of the healthy general population whereas the general population includes sick people along with healthy people.

A few methodological issues might affect the interpretation of our data. First, we used the results of the MINI as a gold standard. The MINI had a good sensitivity and specificity concordance with the Structured Clinical Interview for DSM-IV for diagnosis of MDD (Sensitivity = 0.96, Specificity = 0.86) [Sheehan et al., 1997]. This indicates that the result from the MINI could be reasonably used as a gold standard for MDD. Furthermore, Sheehan et al. [1997] indicated high kappa scores (Kappa = 0.84) for MDD using the MINI as the gold standard.

Second, the study may suffer from an information bias. Even though the privacy of this screening information was guaranteed to all participants, there might have been reluctance from some of them to reveal depressive symptoms in fear of some disadvantage of being diagnosed MDD. This would be a cause of underestimating the number of MDD and increasing the number of false negative of MDD. However, since this same fear applied to both the CES-D and the MINI, this reluctance should not have influenced unduly the validity of the CES-D.

Third, the time frame between the CES-D and the MINI is different. The CES-D asks for symptoms experienced during the preceding 1 week and the MINI asks for symptoms over 2 weeks. These different time frames may have borne some influence on the validity of the CES-D compared to the MINI. Since depression is a chronic disease, we do not think that time difference has influenced the results.

Despite these limitations, the response rate was considerably high in the present study and all participants were interviewed. To our knowledge, this is the first study of the validity of CES-D at a workplace. It shows that the CES-D has a good validity when used in a workplace. A cut-off score of 19 is the optimal cut-off score in assessing MDD in a workforce.

REFERENCES

Mini International Neuropsychiatric Interview (MINI): Concordance and causes for discordance with the CIDI. Eur Psychiatry 13:26–34.


