

University of Nebraska - Lincoln

## DigitalCommons@University of Nebraska - Lincoln

---

Bureau of Sociological Research - Faculty  
Publications

Bureau of Sociological Research (BOSR)

---

June 2003

### A comparison of self-reports of distress and affective disorder diagnoses in rheumatoid arthritis: A receiver operator characteristic analysis

Julia McQuillan

*University of Nebraska - Lincoln, [jmcquillan2@Unl.edu](mailto:jmcquillan2@Unl.edu)*

Judith Fifield

*University of Health Center, Farmington, Connecticut*

T. Joseph Sheehan

*University of Health Center, Farmington, Connecticut*

Susan Reisine

*University of Health Center, Farmington, Connecticut*

Howard Tennen

*University of Health Center, Farmington, Connecticut*

*See next page for additional authors*

Follow this and additional works at: <https://digitalcommons.unl.edu/bosrfacpub>



Part of the [Sociology Commons](#)

---

McQuillan, Julia; Fifield, Judith; Sheehan, T. Joseph; Reisine, Susan; Tennen, Howard; Hesselbrock, Victor; and Rothfield, Naomi, "A comparison of self-reports of distress and affective disorder diagnoses in rheumatoid arthritis: A receiver operator characteristic analysis" (2003). *Bureau of Sociological Research - Faculty Publications*. 12.

<https://digitalcommons.unl.edu/bosrfacpub/12>

This Article is brought to you for free and open access by the Bureau of Sociological Research (BOSR) at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Bureau of Sociological Research - Faculty Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

---

**Authors**

Julia McQuillan, Judith Fifield, T. Joseph Sheehan, Susan Reisine, Howard Tennen, Victor Hesselbrock, and Naomi Rothfield

# A comparison of self-reports of distress and affective disorder diagnoses in rheumatoid arthritis: A receiver operator characteristic analysis

Julia McQuillan,<sup>1</sup> Judith Fifield,<sup>2</sup> T. Joseph Sheehan,<sup>2</sup> Susan Reisine,<sup>2</sup>  
Howard Tennen,<sup>2</sup> Victor Hesselbrock,<sup>2</sup> and Naomi Rothfield<sup>2</sup>

<sup>1</sup> University of Nebraska–Lincoln, Lincoln, Nebraska

<sup>2</sup> University of Health Center, Farmington, Connecticut

Corresponding author: Julia McQuillan, 706 Old Father Hall,  
University of Nebraska–Lincoln, Lincoln, NE 68588; [jmcquillan2@unl.edu](mailto:jmcquillan2@unl.edu)

## Abstract

*Objective*—To compare 3 commonly used psychiatric symptom checklists (the Center for Epidemiological Studies Depression Scale [CES-D], the Positive and Negative Affect Schedule, and the Endler Multidimensional Anxiety Scales [EMAS]) to determine their sensitivity, specificity, and ability to discriminate between a disorder (Major Depression [MD], Generalized Anxiety Disorder [GAD], and no disorder). To compare the checklists for their ability to discriminate between type of disorder (MD and GAD). To evaluate the discriminant ability of the subscales, particularly positive affect; whether the somatic items in the CES-D artificially inflate affective scores; and the optimal cut off score for the CES-D.

*Methods*—We compared the 3 scales to diagnostic criterion of MD, GAD, and comorbid disorder using receiver operator characteristic (ROC) and logistic regression analyses. The sample consisted of a national panel of 415 individuals with rheumatoid arthritis (RA).

*Results*—Each of the scales had high sensitivity and specificity (areas under the curve: CES-D = 0.92, negative affect = 0.88, positive affect and EMAS = 0.82). The CES-D, however, demonstrated better sensitivity and specificity than the positive affect and the EMAS, but not the negative affect scale.

*Conclusion*—All 3 self-reports have high combined sensitivity and specificity as measures of affective disorders among RA patients.

**Keywords:** affective disorder, distress, ROC

## Introduction

Elevated levels of emotional distress indicated by depressive symptoms [1–4], diagnoses of depression [5–8], and anxiety [3] are reported consistently across studies of rheumatoid arthritis (RA) patients. Two reasons individuals with RA may be at greater risk for depression than the general population are high levels of pain and/or functional impairment or a common neurobiologic mechanism underlying both depression and RA [9–11]. An important rival hypothesis is that elevated levels are a measurement artifact.

An accurate assessment of the scope of emotional distress is important and researchers exploring a connection between RA and affective disorders need reliable and valid brief scales for screening individuals. Unfortunately, scores

on the most frequently used measures of psychological distress (screening scales) may be inflated for a variety of reasons, calling into question the meaning currently ascribed to the high distress levels seen in RA samples [1, 2, 8, 12]. The problems stem from 3 sources: 1) findings have not been based on large, representative samples of RA patients; 2) many depression and anxiety scales have adequate convergent validity but low discriminant validity; and 3) there may be overlap between emotional distress items and typical RA symptoms, i.e., criterion contamination [13]. There is particular concern that distress detected by the widely used Center for Epidemiological Studies Depression Scale (CES-D) might not be due to anxiety, depression, or concurrence of the two, but rather that it is an artifact of overlap between CES-D somatic items and RA disease severity [1, 2, 11].

To address these limitations and to test the measurement artifact hypothesis, we employed a large, nationally representative sample; used a structured diagnostic interview; used a multidimensional approach to enhance discriminant validity of screening scales suggested by Clark and Watson [14]; and compared the combined sensitivity and specificity of the CES-D with and without the somatic items using receiver operator characteristic (ROC) curves. This study also provides information to determine appropriate cut off scores for major depression (MD) and generalized anxiety disorder (GAD) for the continuous scales commonly used on the RA population [15-18].

The intensity of depressive symptoms among individuals with RA has been widely studied using various indicators, including the CES-D [1, 2, 11], the Beck Depression Inventory (BDI) [19, 20], and the Arthritis Impact Measurement Scales [3, 19, 21-23]. In samples of individuals with RA, and using the conventional cut off for general populations, investigators have found rates of depression ranging from 23% using the BDI to 46% using the CES-D [1]. Although anxiety symptoms among RA patients have been studied less frequently than symptoms of depression, studies that have assessed anxiety report rates comparable to that reported for depression [3, 19].

When individuals with RA report high levels of psychological distress, it is not clear that researchers should be reporting their findings as evidence of depression. Brief self-report scales are economical and often have good convergent validity (indicated by high correlations with other measures of the same construct), but they do not have good discriminant validity (indicated by high correlations with other measures of related but not identical constructs). It is possible, of course, that while the underlying constructs may be conceptually distinct, they may be empirically collinear. As a result, anxiety measures correlate or load highly with depression measures and it is unclear whether the distress that is being measured is due to anxiety, depression, or both [3, 12, 19, 24-26].

The somatic items included in most measures of depression and anxiety leave open the possibility of criterion contamination [13]. This contamination occurs when items that were designed to assess dimensions of depression or anxiety actually reflect aspects of RA. For example, questions about fatigue or difficulty "getting going" can reflect depression, the effects of RA, or both. Callahan et al [27] and Blalock et al [1] found evidence consistent with the idea that the somatic items of the CES-D are elevated among RA patients due to RA symptoms, not greater depression. Rhee et al [11] found some inflation of total CES-D scores in an RA sample attributable to somatic items. Blalock and colleagues [1] recommended that investigators remove 4 CES-D items to reduce the inflation in scores deriving from RA disease severity.

These problems with common self-report screening questionnaires have compromised our understanding of connections between RA and affective disorders. Fortunately, methods are now available to improve the discriminant validity of questionnaire measures to differentiate between depression and anxiety. Two such methods are used

in this study. The first is a measure of state anxiety that was designed specifically to distinguish anxiety from depression [25]. The second approach is based on the assessment of positive affect (PA) and negative affect (NA), as suggested by Clark and Watson [14]. NA refers to feeling "upset or unpleasantly engaged rather than peaceful," and PA refers to feeling a "zest for life and pleasurable engagement" [14]. These 2 affects can be measured as distinct and orthogonal factors [14]. They discriminate between anxiety and depression because NA is associated with both depression and anxiety, but low PA is unique to depression. These findings have been reported in nonclinical [28, 29] and clinical samples [30-32]. In this study, we assess whether the discriminant validity of the CES-D can be increased by using 2 subscales of the CES-D, identified in our previous work [33], that closely resemble PA and NA.

In summary, we pose the following questions: 1) Which of the commonly used distress scales demonstrates the highest sensitivity and specificity? 2) Do subscale scores improve our ability to distinguish between anxiety disorders and depressive disorders, and does the measurement of PA enhance this distinction? 3) Do somatic items in depression scales, such as the CES-D, artificially inflate reports of the prevalence of affective disorder among individuals with RA, and does removing those items improve scale sensitivity and specificity? 4) What is the optimal cut off score on the widely used CES-D to detect affective disorder in a nationally representative RA sample?

## Patients And Methods

**Patients.** A subset of patients from the National Rheumatoid Arthritis Study (NRAS) were recruited for this project. NRAS was a prospective panel study that completed its tenth and final year in 1997-1998. The panel of 988 patients with classic or definite RA [34] was recruited from a national, random sample of board-certified rheumatologists (details of the recruitment are published elsewhere [[35]]. At the close of their eighth year interview, the 508 patients remaining in the panel were asked if they would be interested in participating in an additional interview about their emotional and physical well being. A total of 462 (91%) agreed to the followup interview and 415 (90%) completed it. Consistent with the 3:1 prevalence of RA among middle-aged women to men, 83% of the sample was female. They were largely upper middle-aged (mean  $58 \pm SD 9.7$  years), married (68%), and out of the labor force (65%). The relatively low employment rate is similar to other samples of RA patients [36].

**Methods.** Diagnostic and scale responses were obtained by telephone interviews that lasted approximately 30 minutes. Current and lifetime psychiatric diagnoses of MD, GAD, and comorbid disorder (CD) were obtained using the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA) [37-39] based on the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV)* [40]. The SSAGA is suitable for either telephone or face-to-face administration by lay interviewers [38, 39].

The SSAGA was selected because it provides a complete and detailed lifetime psychiatric history for adults. The SSAGA interview schedule covers the major *Diagnostic and Statistical Manual of Mental Disorders Revised Third Edition*, DSM-IV, *International Classification of Diseases, 10<sup>th</sup> Revision* defined Axis I psychiatric disorders. In a combined sample of subjects drawn from the general population and from outpatient psychiatric patients, the SSAGA has been shown to have good within and between site reliability [37]. Using a combined sample of outpatient psychiatric patients and subjects drawn from the community with unknown psychiatric histories, the SSAGA has been shown to be valid compared with other standardized psychiatric diagnostic instruments, i.e., the Schedule for Clinical Assessment in Neuropsychiatry (SCAN) [38]. The SSAGA is currently being used in more than 55 studies in the US and 10 studies in foreign countries.

The SSAGA is useful in the arthritis population because it links episodes of MD or GAD to comorbidities and flares. Interviewers, supervised by a clinical psychologist, completed approximately 20 hours of SSAGA training. All of the interviews were edited for accuracy by a research staff member with a master's degree in psychology and several years of experience with development of the interview.

**Measures.** *The Center for Epidemiological Studies Depression Scale.* The CES-D scale consists of 20 questions chosen to reflect various aspects of depression, including depressed mood, feelings of guilt and worthlessness, feelings of helplessness and hopelessness, psychomotor retardation, loss of appetite, and sleep disturbance [41]. Respondents are asked to think of the last week and report the frequency of occurrence for each item on the following 4-point scale 0) rarely, that is less than 1 day; 1) some of the time, 1 to 2 days; 2) a moderate amount of the time, 3 to 4 days; or 3) most or all of the time, 5 to 7 days. Scores can range from 0-60. Aneshensel et al [42] found that phone and in person administrations produce comparable scores. Past studies show that the CES-D demonstrates adequate test-retest stability, as well as concurrent validity and construct validity [41].

Although the CES-D appears to suffer from the same discriminant validity problems as other popular depression questionnaires used in arthritis research [1], it has been chosen for this study for 3 reasons. First, along with the BDI, it has been rated among the best self-report measures of depression and anxiety based on content validity [43]. Second, it has been widely used for some time among arthritis researchers, and knowledge about its validity or how to improve it would be very useful in the analysis of data appearing in the literature. Finally, previous analyses have identified the factor structure of the CES-D, which may enhance its discriminant validity [33].

Blalock et al [1] identified 4 somatic CES-D items that could be confounded with RA symptoms (I felt that everything I did was an effort, My sleep was restless, I felt hopeful about the future, I could not get going). We created a version of the CES-D that removed these somatic items and prorated the remaining items to retain the range of the original scale.

*The state scale of the Endler Multidimensional Anxiety Scales (EMAS).* This scale was used to measure state anxiety [44]. The EMAS-State is a 20-item measure with each item rated on a 5-point intensity scale. It includes a 10-item cognitive-worry subscale and a 10-item autonomic-emotional scale. Previous studies report good reliability for both scales (0.84 or better). Fifield [45] has reported that administration by telephone and questionnaire produce comparable results

*The Positive and Negative Affect Schedule (PANAS).* The PANAS [32] was used as the measure of positive and negative affect. The measure includes two 10-item mood scales for NA and PA, which form independent factors. Each item is rated on a 5-point scale from 1 indicating "felt very slightly or not at all" to 5 indicating "felt very much." Watson et al [32] have reported extensively on the reliability and validity of the scales, finding strong convergent and discriminant validity with other lengthier mood measures [32]. We used the time frame that refers to "over the past week" to make the PANAS comparable to the CES-D and the EMAS. Fifield [45] has reported that administration by telephone and questionnaire produce comparable results

*Diagnosis of affective disorder (MD, GAD, CD).* DSM-IV [40] criteria to qualify for a diagnosis of MD include 1) depressed mood or loss of interest and pleasure in things that the individual usually cared about or enjoyed every day or nearly every day for 2 weeks or more at some time in the past while experiencing impaired role functioning; and 2) 4 of 8 additional symptoms: problems with appetite, sleep, fatigue, energy, interest, self worth, cognition, or suicidal ideas. The episode could not be due to injury, illness, medication or alcohol, childbirth, or the loss of a loved one within certain time parameters. A diagnosis of current MD required that symptoms occurred within 3 weeks of the interview, whereas a diagnosis of lifetime MD required that the symptoms occurred anytime in the past but not in the past 3 weeks.

DSM-IV criteria for GAD include 1) excessive anxiety and worry for 6 months; 2) difficulty controlling worry; and 3) at least 3 of the following 6 symptoms experienced nearly every day during the episode: restlessness, being easily fatigued, difficulty concentrating or mind going blank, irritability, muscle tension, sleep disturbance. Symptoms cannot be due to the direct effects of a substance or medical condition, they must manifest themselves in role impairment, and they cannot be exclusively coincident with mood disorder. Current GAD required all of these criteria within the 6 months preceding the interview, whereas lifetime GAD required the same criteria for 6 months in the past prior to that time.

DSM-IV does not include criteria for a CD. Previous studies show that those meeting criteria for both disorders have more severe and persistent emotional symptoms with more handicapped social lives [46, 47]. Therefore, we classified people who met full criteria for both MD and GAD as CD for the preliminary analysis. For the ROC analysis, people with CD are combined with the people diagnosed with MD or GAD in a general affective disorder category ( $n = 37$ ) or, in the analyses attempting to discriminate be-

tween MD and GAD, they are eliminated ( $n = 27$ ) because they can not be placed in either group.

**Statistical analysis.** We use a variety of methods to assess the effectiveness of the 3 scales to discriminate between those with anxiety, depression, both, or neither. Preliminary analyses of convergent and discriminant validity include bivariate correlations among the full and subscales and analysis of variance (ANOVA) tests of the differences in mean scale scores by affective disorder (MD, GAD, CD). We used Scheffe and Bonferroni tests for multiple post hoc comparisons between the groups. ROC curves simultaneously estimate the sensitivity and specificity of the screening scales, and provide significance tests for the differences between scales. ROC analysis has been used in other studies to assess the sensitivity and specificity of screening tests for depression and/or anxiety [15-17, 48]. However, ROC analysis cannot evaluate the combined effect of subscales to discriminate between groups. Interaction terms created by multiplying subscale scores provide this information, but cannot be used in ROC analysis. Therefore, we tested the possibility that the interaction between the PA and depressed affect (DA) subscales of the CES-D and the PA and NA subscales of the PANAS scale provide more information than the full scale or than the subscales in an additive logistic regression model.

The specificity and the sensitivity rates for each scale provide information about the ability of the instruments to discriminate between those with a disorder and those without, and between disorders (MD and GAD). Determining specificity and sensitivity usually requires choosing cut off scores for the continuous scales, therefore making the results dependent on the selected cut off. Somoza and Mossman [49] and Metz [50] suggest ROC analysis as a way to overcome the problem of having several cut off scores or selecting only one. ROC analysis provides an overall description of a scale's combined sensitivity and specificity (ability to discriminate) throughout its entire range of possible cut off scores, summarized in an area score ( $A_z$ ).

ROC analysis creates an estimated, smoothed curve with a confidence interval for generalizing to the population. Generalizing requires assuming that the scale's scores are normally distributed within each of the groups (disorder/no disorder or MD/GAD), even if the means in each group are different (binormal assumption). Swets [51] indicates that the binormal assumption is valid for a wide variety of diagnostic tests, and Hanley [52] describes ROC analysis as robust to deviations from normality. We used the CLABROC program, part of the ROCKIT software [53], to conduct the ROC analysis. CLABROC is a maximum likelihood program that creates estimated curves for correlated, paired, continuous distributions. The output from the CLABROC program provides information to create ROC curves, a confidence interval around the area under the curve ( $A_z$ ), and comparisons between 2 scales on the same group of people. The  $A_z$  parameter indicates how much the scale improves discrimination over chance. The  $A_z$  is an overall index of the accuracy of the scale. Higher  $A_z$  scores indicate greater combined sensitivity and specificity.

Differences in the accuracy of the scales were tested using the area under the curve test (95% confidence level, 2-tailed test). According to Mossman and Somoza [54], the  $A_z$  represents the probability that 1 of the patients who was diagnosed with the scale as having an affective disorder ranked higher on a particular scale than another patient, who was randomly identified and did not have an affective disorder, ranked on the same particular scale.

To use the ROC analysis to select an optimum cut-off score for a scale, prevalence rates need to be specified because discrimination by score varies by prevalence. Because the optimal cut-off scores depend on the prevalence in the population and the research/clinical situation, it is impossible to provide a single best score. However, we calculated the true positive rate and the false positive rate for various cut off scores for all of the scales (available from the first author) [16].

Logistic regression analysis provides a way to assess the effectiveness of combinations of the subscales (interaction terms) to discriminate between those with and those without an affective disorder, or between disorders. Interaction terms that combine information from the PA and NA subscales of the PANAS and the PA and DA subscales of CES-D should have larger coefficients than either scale alone if Clark and Watson's [14] tripartite model of affective disorder is correct. However, a problem arises when trying to interpret odds ratios when the independent variables are continuous. Because the coefficients describe changes per unit change in the independent variable, for the coefficients to be comparable, the independent variables need to be standardized by subtracting the scale mean from each value of the independent variable and dividing by the standard deviation of the independent variable.

## Results

**Sample description and prevalence of affective disorders.** Similar to the RA population, most of the study participants are women (83%) and married (68%), but only one-third are currently employed (35%). At entry into the study, participants had been living with a diagnosis of RA for an average of 10 years; the mean age at diagnosis was 39.5 years. Median family income was \$35,000 and median education was 1 year of post-high school education. Average pain and fatigue levels in the sample during the eighth wave of the study were at the middle of the range (45 for pain, 51 for fatigue). Of the 415 participants in this study, 9% met the criteria for current major depressive episode (MD = 4%), current generalized anxiety disorder (GAD = 3%), or both simultaneously (CD = 2%).

**Means and alpha reliabilities of the scales.** All of the screening scales had adequate alpha reliabilities (Table 1). The lowest was 0.71 for the interpersonal subscale of the CES-D, the highest is 0.93 for the EMAS summary scale. We also constructed a CES-D scale with the somatic items removed (CES-D<sub>nos</sub>). The items for this scale were prorated to make a scale with the same possible range of scores as the original CES-D scale. The mean for the scale without the somatic items (mean = 10.17) is lower than for

the scale with the somatic items (mean = 12.23, paired *t*-test  $P < 0.05$ ); however, the 2 scales were almost perfectly correlated (0.99). Squaring this coefficient ( $R^2 = 0.97$ ) indicates that less than 3% of the original CES-D score is explained by the somatic items. Although the scores were significantly lower when the somatic items were removed, suggesting some RA contamination, the magnitude of the difference is small. We further explore how much the somatic items inflate the CES-D scores among people with RA using the ROC analyses.

**Correlations among the scales and subscales.** Table 2 provides the correlations among the scales and subscales for all of the cases ( $n = 415$ , bottom half of the diagonal) and for just those cases with an affective disorder ( $n = 37$ , top half of the diagonal).

We focus first on the correlations in the entire sample. The first 5 columns contain the subscales of the CES-D, as identified by Sheehan et al [33]. All of the correlations are above 0.60, indicating adequate convergent validity. All of the subscales had strong positive correlations with the full scale (0.80-0.93). These patterns of correlations provide evidence of good convergent validity for the CES-D subscales.

Next, we assessed the correlations between the subscales of the CES-D and the subscales of the EMAS and the PANAS. Some of the correlations between the depression subscales and the anxiety subscales are quite high (5 correlations are higher than 0.70, the lowest was 0.43). The strong positive correlation between both the depression subscales (CES-D) and the anxiety subscales (EMAS) and NA indicates that both scales were tapping negative affect. NA and PA had, as they should, a modest correlation ( $r = 0.36$ ). The correlations for study participants with a diagnosis of either MD or GAD ( $n = 27$ ) are similar to the full sample. The smallest correlation is -0.05, the largest is 0.86. The overall pattern of correlations among those with an affective disorder is similar to the pattern among all study participants; there is better convergent validity than discriminant validity.

**Table 1.** Means and reliabilities for the CES-D, EMAS, and PANAS with subscales ( $n = 415$ )<sup>\*</sup>

Scale	Mean	SD	Alpha reliability	Min-max
CES-D	12.23	11.00	0.75	0-50
CES-Dnos0	10.17	10.73	0.72	0-47
CES-D subscales				
Somatic	5.79	4.44	0.83	0-21
Interpersonal	0.96	1.68	0.71	0-10
Depressive	2.66	3.37	0.88	0-15
Positive	9.19	3.02	0.83	0-12
EMAS	29.83	12.05	0.93	20-84
Autonomic worry	14.50	5.76	0.83	10-40
Cognitive worry	15.21	6.95	0.91	10-40
PANAS positive affect scale	30.49	7.98	0.89	10-50
PANAS negative affect scale	17.69	7.48	0.91	10-50

<sup>\*</sup> CES-D = Center for Epidemiological Studies Depression Scale; EMAS = Endler Multidimensional Anxiety Scales; PANAS = Positive and Negative Affect Schedule; SD = standard deviation; Min = minimum; max = maximum; CES-Dnos0 = CES-D with somatic items removed.

**Distinguishing between MD, GAD, and CD.** We compared mean CES-D, NA, PA, and EMAS scores for those with no affective disorder ( $n = 378$ ), those with MD ( $n = 16$ ), GAD ( $n = 11$ ), and CD ( $n = 10$ ) using ANOVA (Table 3).

The group without an affective disorder had lower average scores on all of the scales than the groups with a disorder. The average scores for those in the affective disorder categories were at least 1 standard deviation above the mean for all scales. These results indicate that each of the screening questionnaires were able to distinguish between individuals with an affective disorder and those without one. There were large overlaps in the confidence intervals between types of disorders (GAD, MD, or CD), suggesting that none of the scales were able to differentiate between the types of disorders. Furthermore, the subscales showed no apparent advantage over the full CES-D or EMAS. PA scores were comparable for those with GAD, MD, or both.

**Table 2.** Correlations among the scales and subscales<sup>\*</sup>

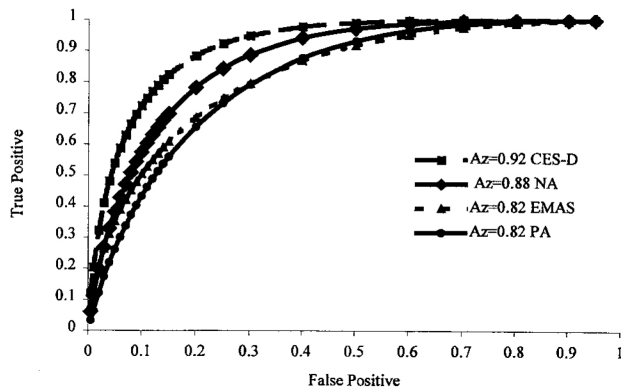
	1	2	3	4	5	6	7	8	9	10
1 CES-D interpersonal		0.46	0.56	0.51	0.76	0.48	0.35	0.45	<b>0.25</b>	0.46
2 CES-D somatic vegetative	0.63		0.70	<b>0.32</b>	0.85	0.53	0.34	0.47	0.22	0.41
3 CES-D depressed affect	0.73	0.79		0.34	0.86	0.47	<b>0.27</b>	0.40	0.33	0.51
4 CES-D positive affect (reversed)	0.60	0.61	0.68		0.64	0.38	<b>0.08</b>	<b>0.25</b>	0.41	0.45
5 CES-D summary score	0.80	0.91	0.93	0.82		0.59	0.34	0.50	0.37	0.57
6 EMAS cognitive worry	0.65	0.73	0.70	0.59	0.77		0.72	0.93	<b>0.09</b>	0.53
7 EMAS autonomic worry	0.52	0.65	0.57	0.43	0.63	0.75		0.92	<b>-0.20</b>	0.30
8 EMAS summary score	0.63	0.74	0.69	0.56	0.76	0.95	0.92		<b>-0.05</b>	0.45
9 Positive affect (reversed)	0.41	0.49	0.47	0.63	0.58	0.41	0.30	0.38		<b>0.10</b>
10 Negative affect	0.67	0.72	0.75	0.60	0.79	0.76	0.64	0.75	0.36	

<sup>\*</sup> Bottom half contains all study participants ( $n = 415$ ), top half (italics) contains just those with an affective disorder ( $n = 37$ ). All of the correlations in the bottom left half are significant at the 0.01 level (2-tailed) ( $n = 415$ ). All of the correlations in the top right half are significant at the 0.05 level (2-tailed) except for those in bold ( $P$  values are greater than 0.05) ( $n = 37$ ). See Table 1 for definitions.

**Table 3.** Means, SDs, and 95% CIs for the scales by affective disorder category\*

Scale	Affective disorder category	N	Mean	SD	95% CI
Positive affect	No disorder	378	31.30	7.62	30.53-32.07
	MD	16	22.81	7.53	18.80-26.82
	GAD	11	22.18	6.66	17.71-26.66
	CD	10	21.40	6.69	16.62-26.18
Negative affect	No disorder	378	16.66	6.62	15.99-17.33
	MD	16	27.25	7.61	23.19-31.31
	GAD	11	28.73	9.26	22.50-34.95
	CD	10	29.10	6.74	24.28-33.92
EMAS	No disorder	378	28.40	10.64	27.33-29.48
	MD	16	47.94	16.84	38.96-56.91
	GAD	11	41.09	16.08	30.29-51.89
	CD	10	42.30	13.74	32.47-52.13
CES-D	No disorder	378	10.48	9.38	9.53-11.42
	MD	16	30.94	10.00	25.61-36.26
	GAD	11	27.27	10.07	20.51-34.04
	CD	10	32.20	11.37	24.07-40.33
CES-Dnoso	No disorder	378	8.46	9.09	7.54-9.38
	MD	16	27.97	10.06	22.61-33.33
	GAD	10	24.50	10.66	16.88-32.12
	CD	10	29.63	12.15	20.93-38.32

\* SDs = standard deviations; 95% CIs = 95% confidence intervals; MD = major depression; GAD = generalized anxiety disorder; CD = comorbid disorder. See Table 1 for other definitions.



**Figure 1.** Receiver operator characteristic curves for the full scales (Center for Epidemiological Studies Depression Scale [CES-D], negative affect [NA], Endler Multidimensional Anxiety Scales [EMAS], and positive affect [PA]).

**ROC analysis.** We first used ROC analysis to test the ability of the scales and subscales to discriminate between those with GAD and those with MD (n = 27). Because the CD participants fit in neither group, they are omitted from this analysis and only are incorporated in the analysis of those with and those without a disorder. All of the scales showed poor discriminant validity (Az scores range from 0.45 to 0.66; none were significantly better than chance). Only 2 of the subscales had Az scores that were different from their full scales: the interpersonal and DA subscales of the CES-D were each significantly *worse* than the full CES-D at discriminating between depression and anxiety.

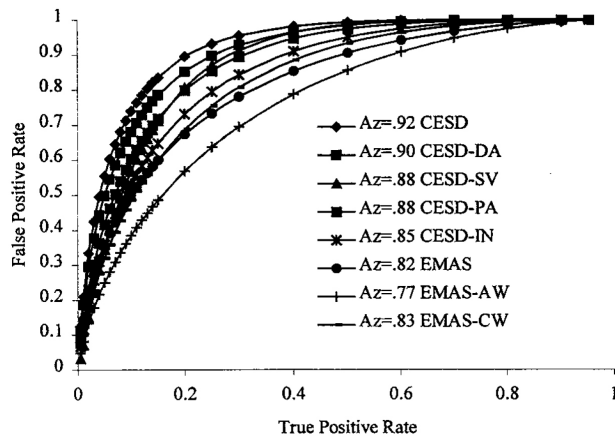
Figure 1 shows the ROC analyses of the full scales using affective disorder (AD), that is MD or GAD or both, as the criterion. The ROC curves for the CES-D, NA, EMAS, and PA show areas under the curve ranging from a high of

0.92 for the CES-D to 0.82 for the PA. The CES-D had a significantly higher area under the curve than the EMAS (Az = 0.92 versus 0.82; *P* = 0.003) and the PA (Az = 0.92 versus 0.82; *P* = 0.003), but not the NA. These results indicate that the CES-D is better than the EMAS, PA, or NA at differentiating between those with and those without an affective disorder. Although not shown in the figure, the curve for the CES-D without the somatic items was not different from the CES-D with the somatic items.

All of the Azs were high, suggesting that all of the scales had good combined sensitivity and specificity (all are higher than 0.80) to predict membership in the affective disorder group compared with the nonaffective disorder group. We also tested the subscales of the CES-D and the EMAS compared with the full scales to ascertain if they added information hidden in the combined scales (Figure 2). None of the subscales had significantly higher Az scores than the full scales. The interpersonal, DA, and PA subscales had significantly lower Az scores than the full CES-D scale.

**Logistic regression analysis.** We tested the possibility that the interaction between the PA and DA subscales of the CES-D and the PA and NA subscales of the PANAS scale provide more information than either full scale or than the subscales in an additive logistic regression model. The EMAS subscales have not previously been examined for interactive effects; however, we decided to test the possibility that the subscales might be more effective than the full scale in a fashion parallel to the CES-D and PANAS. We focused only on the whole sample (n = 415) because the disordered subsample did not have sufficient cases for the analysis (the Hoesmer-Lemeshow test showed that several categories had fewer than the minimum required 5 expected cases).





**Figure 2.** Receiver operator characteristic curves for the full and subscales of the Center for Epidemiological Studies Depression Scale (CESD) and Endler Multidimensional Anxiety Scales (EMAS) ( $n = 415$ ). CESD-DA = CESD depressed affect; CESD-SV = CESD somatic vegetative; CESD-PA = CESD positive affect; CESD-IN = CESD interpersonal; EMAS-AW = EMAS autonomic worry; EMAS-CW = EMAS cognitive worry.

The logistic regression analysis revealed that both the full CES-D and the PA and DA subscales of the CES-D showed significant positive associations with the likelihood of having an affective disorder diagnosis. Combining information from the DA and PA subscales does not contribute to the simple additive model. The pattern is the same for the PANAS, but slightly different for the EMAS. For the EMAS, when both of the subscales (autonomic worry and cognitive worry) were in the model, the autonomic worry subscale ceased to be a significant predictor, suggesting that there was considerable overlap between these subscales.

**Cut point for the CES-D.** ROC analysis also allowed us to evaluate the best cut point for determining membership in the affective disorder category. The generally accepted cut point for the CES-D is 16. Some researchers have questioned the applicability of this value for the RA population because of the possible criterion contamination of the somatic items. We saw little evidence of contamination. Both the false positives and the true positives increased substantially when the cut point scores were increased from 17.5 to 20.5. There was very little difference between scores of 16.5 and 17.5.

The ROC program automatically selects cut point scores from the raw data. Because cut points of 16 and 19 have been suggested in past research, we examined these 2 cut point scores for the full CES-D and the CES-D<sub>nos</sub> (somatic items removed) using simple cross-tabulated data (Table 4). In this sample, there was a difference of only 1 “true” case missed when 19 was used as a cut score, compared with 16, but there were 22 more false positives with 16. Like other studies using the CES-D, we found that the specificity (0.89) of the CES-D is much better than the sensitivity (0.24) at the standard cut off score of 16.

## Discussion

This study provided a unique opportunity to review the usefulness of 3 popular scales of depression and anxiety for use as screeners in RA research. Its uniqueness stems from the national representative sample and the availability of both screener and diagnostic interview scores. Using these data, we found that it is possible to detect affective disorders among people with RA using short symptom scales. Furthermore, we found much lower rates of depression and anxiety in this sample than reported in other RA studies; the percentages are closer to the national average than previously found in RA samples [55]. We do not attribute this lower rate to our decision to exclude dysthymia from our analyses, because only an additional 2% of the sample (8 cases) met criteria for dysthymia only. Rather, we conclude that higher rates of depression and depressive symptoms found in other studies are more likely due to the characteristics of the samples than weaknesses in the scales, because the scales, particularly the CES-D, demonstrated high combined sensitivity and specificity.

Similar to Breslau [12], we found that the CES-D does not differentiate between MD and GAD, but does detect high levels of depression and anxiety equally well. None of the self-report measures of distress discriminated between DSM-IV MD and GAD. Despite the promise of Clark and Watson’s [14] tripartite model for differentiating between anxiety and depression using positive and negative affect subscales, PA did not differentiate between MD and GAD in this sample. We do not have a satisfactory explanation for this result. Possible explanations include sample-specific findings, a small number of individuals with MD or GAD, something particular about the RA population, or that MD and GAD share too much of a common core to be discernible with general scales. The evidence of comorbidity in this and other studies suggests that differentiating between MD and GAD may be less important than identifying the existence of an affective disorder.

**Table 4.** Comparing the CES-D and the CES-D<sub>nos</sub> at 2 conventional cut points ( $n = 415$ )\*

Scale	Cut point	n (%)	True positives missed	False positives	Sensitivity	Specificity
CES-D	16	122 (29)	4	89	0.89	0.24
CES-D <sub>nos</sub>	16	114 (27)	4	81	0.89	0.21
CES-D	19	99 (24)	5	67	0.86	0.18
CES-D <sub>nos</sub>	19	86 (21)	8	57	0.78	0.15

\* CES-D = Center for Epidemiological Studies Depression Scale; CES-D<sub>nos</sub> = CES-D with somatic items removed.

Because the consequences of comorbidity are quite high, physicians should attempt to detect and treat those cases that emerge. In addition to RA-affective disorder comorbidity, there is evidence of comorbidity among anxiety disorders. Fifty to sixty percent of individuals who meet criteria for 1 anxiety disorder diagnosis also meet criteria for additional comorbid anxiety diagnoses [56]. Individuals with anxiety disorders are also likely to have major depression and substance use disorders [57]. When they occur together, anxiety disorders tend to precede the development of mood disorders, such as depression or substance use [58]. This temporal ordering suggests that anxiety may contribute to the development of other psychiatric problems, making early detection a high priority.

What do these results suggest for clinical practice and future research on RA/affective disorder connections? In clinical practices where the costs of false positives are low, using a cut point of 16 for the CES-D with followup screening seems prudent. In prevalence research where questions about the relationship between RA and affective disorders are the focus, a cut point of 19 for the CES-D is more conservative and therefore preferable. If interview time is limited, using the DA subscale of the CES-D is almost as reliable as the full scale; however, cut points for the subscale need to be established.

Researchers should be cautious in assuming a higher prevalence of MD or GAD in RA populations. This study, using a representative sample and diagnostic interview schedule derived diagnoses, shows similar rates to a general population national survey. Despite similar numbers, because the consequences of comorbidity are greater than each disorder alone, it is essential to identify those meeting criteria for both affective disorders, and to continue to explore connections between chronic illnesses such as RA and affective disorders.

## References

- Blalock SJ, DeVellis RF, Brown GK, Wallston KA. Validity of the Center for Epidemiological Studies Depression Scale in arthritis populations. *Arthritis Rheum* 1989; **32**: 991-7.
- Fifield J, Reisine S. Characterizing the meaning of psychological distress in rheumatoid arthritis. *Arthritis Care Res* 1992; **5**: 184-91.
- Hawley DJ, Wolfe F. Anxiety and depression in patients with rheumatoid arthritis: a prospective study of 400 patients. *J Rheumatol* 1988; **15**: 932-41.
- Katz PP, Yelin EH. Prevalence and correlates of depressive symptoms among persons with rheumatoid arthritis. *J Rheumatol* 1993; **20**: 790-6.
- Zathropoulos G, Barry H. Depression in rheumatoid disease. *Ann Rheum Dis* 1974; **33**: 132-5.
- Rimon R. Depression in rheumatoid arthritis. *Ann Clin Res* 1974; **6**: 171-5.
- Rimon R, Laakso R. Overt psychopathology in rheumatoid arthritis: a fifteen year follow-up study. *Scand J Rheumatol* 1984; **13**: 324-8.
- Frank RG, Beck NC, Parker JC, Kashani JH, Elliott TR, Haut AE, et al. Depression in rheumatoid arthritis. *J Rheumatol* 1988; **15**: 920-31.
- Sternberg E, Chrousos GP, Wilder RL, Gold PW. The stress response and the regulation of inflammatory disease. *Ann Intern Med* 1992; **117**: 854-66.
- Rechlin T, Weis M, Kaschka W. Is diurnal variation of mood associated with parasympathetic activity? *J Affect Disord* 1995; **34**: 249-55.
- Rhee SH, Petroski GF, Parker JC, Smarr KL, Wright GE, Multon KD, et al. A confirmatory factor analysis of the Center for Epidemiologic Studies Depression Scale in rheumatoid arthritis patients: additional evidence for a four-factor model. *Arthritis Care Res* 1999; **12**: 392-400.
- Breslau N. Depressive symptoms, major depression, and generalized anxiety: a comparison of self-reports on CES-D and results from diagnostic interviews. *Psychiatry Res* 1985; **15**: 219-29.
- Pincus T, Callahan LF. Depression scales in rheumatoid arthritis: criterion contamination in interpretation of patient responses. *Patient Educ Couns* 1993; **20**: 133-43.
- Clark L, Watson D. Tripartate model of anxiety and depression: psychometric evidence and taxonomic implications. *J Abnorm Child Psychol* 1991; **100**: 316-36.
- Somoza E, Mossman D. "Biological markers" and psychiatric diagnosis: risk-benefit balancing using ROC analysis. *Biol Psychiatry* 1991; **29**: 811-26.
- Somoza E, Steer RA, Beck AT, Clark DA. Differentiating major depression and panic disorders by self-report and clinical rating scales: ROC analysis and information theory. *Behav Res Ther* 1994; **32**: 771-82.
- Weinstein MC, Berwich DM, Goldman PA, Murphy JM, Barsky AJ. A comparison of three psychiatric screening tests using receiver operating characteristic (ROC) analysis. *Med Care* 1989; **27**: 593-607.
- Wyshak G, Barsky AJ. Relationship between patient self-ratings and physician ratings of general health, depression, and anxiety. *Arch Fam Med* 1994; **3**: 419-24.
- Hagglund KJ, Roth DL, Haley WE, Alarcon GS. Discriminant and convergent validity of self-report measures of affective distress in patients with rheumatoid arthritis. *J Rheumatol* 1989; **16**: 1428-32.
- Smith TW, Peck JR, Ward JR. Helplessness and depression in rheumatoid arthritis. *Health Psychol* 1990; **9**: 377-89.
- Parker J, Smarr K, Anderson S, Hewett J, Walker S, Bridges A, et al. Relationship of changes in helplessness and depression to disease activity in rheumatoid arthritis. *J Rheumatol* 1992; **19**: 1901-5.
- Parker JC, Wright GE. The implications of depression for pain and disability in rheumatoid arthritis. *Arthritis Care Res* 1995; **8**: 279-83.
- Pincus T, et al. Elevated MMPI scores for hypochondriasis, depression, and hysteria in patients with rheumatoid arthritis reflect disease rather than psychological status. *Arthritis Rheum* 1986; **29**: 1456-66.
- Gotlib I. Depression and general psychopathology in university students. *J Abnorm Psychol* 1984; **93**: 19-30.
- Endler N, Cox BJ, Parker JD, Bagby RM. Self-reports of depression and state-trait anxiety: evidence for differential assessment. *J Pers Soc Psychol* 1992; **63**: 832-8.

- 26 Tanaka-Matsumi J, Kameoka VA. Reliabilities and concurrent validities of popular self-report measures of depression, anxiety, and social desirability. *J Consult Clin Psychol* 1986; **54**: 328-33.
- 27 Callahan LF, Kaplan MR, Pincus T. The Beck Depression Inventory, Center for Epidemiological Studies Depression Scale (CES-D), and General Well-Being Schedule depression subscale in rheumatoid arthritis: criterion contamination of responses. *Arthritis Care Res* 1991; **4**: 3-11.
- 28 Tellegen A. Structures of mood and personality and their relevance to assessing anxiety with an emphasis on self-report. In: Tuma A, Maser JD, editors. *Anxiety and the anxiety disorders*. Hillsdale (NJ): Erlbaum; 1985. p. 681-706.
- 29 Watson D, Tellegen A. Toward a consensual structure of mood. *Psychol Bull* 1985; **98**: 219-35.
- 30 Bouman TK, Luteijn F. Relations between the pleasant events schedule, depression, and other aspects of psychopathology. *J Abnorm Psychol* 1986; **95**: 373-7.
- 31 Hall C. Differential relationships of pleasure and distress with depression and anxiety over a past, present and future time framework [dissertation]. Minneapolis: University of Minnesota; 1977.
- 32 Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. *J Personal Soc Psychol* 1988; **54**: 1063-70.
- 33 Sheehan TJ, Fifield J, Reisine S, Tennen H. The measurement structure of the Center for Epidemiologic Studies Depression Scale. *J Pers Assess* 1995; **64**: 507-21.
- 34 Ropes M, Bennett G, Cobb S, Jacox R, Jessar R. 1958 revision of diagnostic criteria for rheumatoid arthritis. *Bull Rheum Dis* 1958; **9**: 175-6.
- 35 Reisine S, Fifield J, Winkelman DK. Characteristics of rheumatoid arthritis patients: who participates in long-term research and who drops out? *Arthritis Care Res* 2000; **13**: 3-10.
- 36 Yelin E, Henke C, Epstein W. The work dynamics of the person with rheumatoid arthritis. *Arthritis Rheum* 1987; **30**: 507-12.
- 37 Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger JI Jr, et al. A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. *J Stud Alcohol* 1994; **55**: 149-58.
- 38 Hesselbrock MN, Easton C, Bucholz KK, Schuckit M, Hesselbrock V. A validity study of the SSAGA - a comparison with the SCAN. *Addiction* 1999; **94**: 1361-70.
- 39 Hesselbrock MN, Hesselbrock VM, Tennen H, Meyer RE, Workman KL. Methodological considerations in the assessment of depression in alcoholics. *J Consult Clin Psychol* 1983; **51**: 399-405.
- 40 *Diagnostic and Statistical Manual of Mental Disorders*. Washington (DC): American Psychiatric Association; 1994.
- 41 Radloff L. The CES-D scale: a self report depression scale for research in the general population. *Appl Psychol Measure* 1977; **1**: 385-401.
- 42 Aneshensel CS, Frerichs RR, Clark VA, Yokopenic PA. Measuring depression in the community: a comparison of telephone and personal interviews. *Public Opin Q* 1982; **46**: 110-21.
- 43 Gotlib I, Cane D. Self-report assessment of depression and anxiety. In: Kendall PC, Watson D, editors. *Anxiety and depression: distinctive and overlapping features*. New York: Academic Press; 1989. p. 131-69.
- 44 Endler N, Edwards J, Vitelli R. *Endler Multidimensional Anxiety Scales (EMAS): manual*. Los Angeles: Western Psychological Services; 1991.
- 45 Fifield J, Tennen H, Reisine S, Hesselbrock V, Sheehan TJ, McQuillan J, et al. A comparison of risk factors for major depression and depressive symptoms reports in rheumatoid arthritis. Arthritis Research Conference. Bethesda (MD), NIAMS, May 29-June 1 1997.
- 46 Hecht H, von Zerssen D. Anxiety and depression in a community sample: the influence of comorbidity on social functioning. *J Affect Disord* 1990; **18**: 137-44.
- 47 Newman SC, Bland RC. Life events and the 1-year prevalence of major depressive episode, generalized anxiety disorder, and panic disorder in a community sample. *Compr Psychiatry* 1994; **35**: 76-82.
- 48 Wyshak G, Barsky AJ, Klerman GL. Comparison of psychiatric screening tests in a general medical setting using ROC analysis. *Med Care* 1991; **29**: 775-85.
- 49 Somoza E, Mossman D. ROC curves and the binormal assumption. *J Neuropsychiatry Clin Neurosci* 1991; **3**: 436-9.
- 50 Metz CE. Basic principles of ROC analysis. *Semin Nucl Med* 1978; **8**: 283-98.
- 51 Swets JA. ROC analysis applied to the evaluation of medical imaging techniques. *Invest Radiol* 1979; **14**: 109-21.
- 52 Hanley JA. The robustness of the "binormal" assumptions used in fitting ROC curves. *Med Decis Making* 1988; **8**: 197-203.
- 53 Metz CE, Herman BA, Roe CA. Statistical comparison of two ROC-curve estimates obtained from partially-paired datasets. *Med Decis Making* 1998; **18**: 110-21.
- 54 Mossman D, Somoza E. ROC curves, test accuracy, and the description of diagnostic tests. *J Neuropsych Clin Neurosci* 1991; **3**: 330-3.
- 55 Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, et al. Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: results from the National Comorbidity Survey. *Arch Gen Psychiatry* 1994; **51**: 8-19.
- 56 Brown TA, Barlow DH. Comorbidity among anxiety disorders: implications for treatment and DSM-IV. *J Consult Clin Psychol* 1992; **60**: 835-44.
- 57 Angst J, Vollrath M, Merikangas KR, Ernst C. Comorbidity of anxiety and depression in the Zurich Cohort Study of Young Adults. In: Maser JD, Cloninger CR, editors. *Comorbidity of mood and anxiety disorders*. Washington (DC): American Psychiatric Press; 1990.
- 58 Hesselbrock MN, Meyer RE, Keener JJ. Psychopathology in hospitalized alcoholics. *Arch Gen Psych* 1985; **42**: 1050-5.