Measurement of Depressive Symptoms in Women With Breast Cancer and Women With Clinical Depression: A Differential Item Functioning Analysis

Niels G. Waller,1,4 Bruce E. Compas,1 Steven D. Hollon,3 and Ellen Beckjord1,2

Differential item functioning (DIF) analyses of the Beck Depression Inventory-II (BDI-II) were conducted on samples of 267 women with breast cancer and 294 women with clinical depression. Patterns of items in which there was significant and nonsignificant DIF were identified using statistical tests and measures of DIF effect size. At the most general level, 15 of 21 BDI-II items were associated with nontrivial DIF suggesting that the item responses of these samples do not reflect the same underlying construct. Factor analyses of the BDI-II using a psychometrically defensible method for item level factor analysis supported the conclusions from the DIF analyses. These findings suggest that researchers and practitioners should apply caution when interpreting self-report depression symptoms in breast cancer patients.

KEY WORDS: breast cancer; depression; measurement.

Accurate assessment of depression and depressive symptoms in medically ill patients is of paramount importance. Studies have shown that subclinical symptoms of depression and diagnoses of major depressive disorders are implicated in the psychological and physical health of patients with a wide range of conditions. This includes patients recovering from acute events such as myocardial infarction (e.g., Frasure-Smith, Lesperance, Juneau, Tlarijic, & Bourassa, 1999), and patients suffering from chronic conditions such as HIV infection and AIDS (Rosenberg et al., 2001) and cancer (McDaniel, Musselman, Porter, Reed, & Nemeroff, 1995). Depression is a possible consequence of the stress associated with some medical conditions, is a predictor of course and recovery, and may be an important source of comorbidity that can affect response to treatment.

In spite of its potential importance, the measurement of depression and depressive symptoms associated with disease has been hindered by several factors. Foremost among these problems is the possible confounding of symptoms of depression with symptoms of some diseases and with possible side effects of treatment. These concerns are particularly important in the assessment of depression in cancer patients, as some symptoms of some types of cancer and the side effects of adjuvant therapies can mimic symptoms of depression (Croyle & Rowland, 2003; Raison & Miller, 2003). For example, some forms of chemotherapy are associated with increased fatigue, loss of energy, decreased appetite, loss of sexual drive, and impairment in concentration and attention, all of which are focal symptoms of depression. Disentangling symptoms that are attributable to depression as contrasted with those that are due to disease or side effects of treatment are particularly difficult when depressive symptoms are measured with self-report questionnaires, such as the BDI-II (e.g., the Beck Depression Inventory-II; Beck, Steer, & Brown, 1996; see Dozois & Covin, in press, for a scholarly review). In both

1Department of Psychology and Human Development, Vanderbilt University, Nashville, Tennessee.
2Present address: Division of Cancer Prevention, National Cancer Institute, National Institute’s of Health, Department of Health and Human Services, Bethesda, MD.
3Department of Psychology, Vanderbilt University, Nashville, Tennessee.
4Correspondence should be addressed to Niels Waller, Department of Psychology and Human Development, Peabody College, Vanderbilt University, Nashville, Tennessee 37203; e-mail: niels.waller@vanderbilt.edu.
research and clinical practice, self-report inventories are typically examined only in terms of their aggregate scores and little or no attention is given to other factors that could lead to endorsement of some symptoms as a result of disease or treatment processes.

Breast cancer patients are an optimal population in which to examine factors that may affect depressive symptoms. Breast cancer is the most commonly diagnosed cancer among women in the U.S. (American Cancer Society, 2003), and it has been the focus of extensive and intensive research on the psychological correlates of the disease and its treatment (Compas & Leucken, 2002). However, few studies (Ritterband & Spielberger, 2001) have directly compared the rates and patterns of endorsement of depressive symptoms among women with breast cancer, women free of disease, and women with clinical depression.

The few studies that have been conducted in this area (summarized in Spijker, Trijsburg, & Duivenvoorder, 1997) typically find that cancer patients produce elevated ratings on self-report depression scales when compared to healthy controls. Unfortunately, this finding is difficult to interpret because many researchers have not considered whether depression scales measure depression—and only depression—in cancer patients. Asking this logically prior question, Ritterband and Spielberger (2001) recently noted that “depression measures usually include items that assess somatic and performance difficulties that may not be just symptoms of depression, but rather consequences of disease treatment . . .” Therefore, the total scores of cancer patients on measures of depression may overestimate the severity of depression, resulting in many false positive findings” (p. 86). These remarks point to the need for focusing our analyses on a structural level that is lower than that provided by scale total scores. The most obvious lower level considers item response data.

Potentially, item-level analyses can uncover population differences in scale structure and construct composition that are hidden from subscale or total scale analyses. This could occur, for instance, if the aggregate scores failed to measure similar dimensions in different populations. Ultimately, multivariate techniques such as factor analysis can elucidate scale and item dimensionality in samples that are homogenous with respect to diagnosis, but these analyses must be conducted with appropriate psychometric models. To our knowledge, no study has explored the BDI-II factor structure in cancer patients using psychometrically optimal methods for ordered categorical items (e.g., binary or Likert items, see Bock, Gibbons, & Muraki, 1988; Waller, 2003).

A second desiderata of item level analyses is the ability to explore trait-by-group interactions via models of differential item functioning (DIF; Camilli & Shepard, 1994; Holland & Wainer, 1993; Thissen, Steinberg, & Gerrard, 1986). Recent DIF analyses of the BDI (Kim, Pilkonis, Frank, Thase, & Reynolds, 2002; Santor, Ramsay, & Zuroff, 1994a) and other depression measures (Santor & Coyne, 2001; Santor, Zuroff, Cervantes, Palacios, & Ramsay, 1995) have advanced our understanding of depression assessment in various populations. Thus, a DIF analysis of the BDI-II in women with breast cancer is likely to enhance our understanding of depression assessment in this important population (American Cancer Society, 2003).

In the remainder of this paper we define DIF, describe how to measure DIF with standard statistical software, and then perform a DIF analysis of the BDI-II in women with breast cancer and women with clinical depression. We then explore the factor structure of the BDI-II using a psychometrically defensible method for item level factor analysis (Knol & Berger, 1991; Waller, 2003) and a rotation method (Schmid & Leiman, 1957) that simultaneously elucidates the general and group factors of the inventory. We conclude by discussing the implications of our results for depression assessment in women with breast cancer.

Measuring Differential Item Functioning With Logistic Regression

Stated plainly, a DIF study asks the following question: For trait levels, θ, does the probability (P) of an item response (U) for an individual from Group A differ from that of an individual from Group B? Notice in this definition that we are not comparing group means (e.g., by t-tests) or group item response rates (by χ² tests). On the contrary, all group comparisons are made with trait levels held constant. This is an important feature of DIF methodology that allows DIF studies to identify biased items.

An item is biased if it produces different item-trait regression functions in different groups. To understand this statement, consider the definition of an unbiased item. If an item is unbiased—that is, if an item shows no DIF—then the conditioned response probabilities can be expressed:

\[ P_j(U | \theta, g = A) = P_j(U | \theta, g = B), \] (1)

where \( g \) is a group designator, \( j \) is an item index, and all other terms are defined as above. In plain English, this equation states that the probability that an individual