

## Dementia Caregiver Depression: An Analysis of Meaning and Symptom Severity

La Tonya Noël\*, Robert L. Glueckauf, Megan Deichen

Florida State University, College of Social Work, FL, USA

\*Corresponding author: Dr. La Tonya Noël, Ph.D., Florida State University, College of Social Work, PO BOX 3062570, UCC 2512, Tallahassee, FL, 32306-2570, Tel: 850-645-1571; Email: lnoel@fsu.edu

Received: 03-09-2015

Accepted: 10-13-2015

Published: 11-24-2015

Copyright: © 2015 Noël

### Abstract

This study explored differential symptom endorsement on the Patient Health Questionnaire-9 (PHQ-9), associated item severity ratings, and assessed the impact of item complexity on caregiver responses between two samples of family dementia caregivers: an African-American (n = 8) and a non-Hispanic White comparison group (n = 6). Semi-structured interviews were performed to explore patterns of endorsement and symptom interpretation. The two groups responded in substantially different ways on three of the nine PHQ-9 items. The implications of both the qualitative and quantitative findings on assessing caregiver depression are discussed. Future directions for research on evaluation of family caregiver depression and instrument use/restructuring are also addressed.

**Keywords:** Caregiving; Dementia; Depression; Symptom Endorsement; PHQ-9

### Introduction

Providing regular assistance for a loved one with progressive dementia, such as Alzheimer's disease, has been associated with perceived emotional and social burden and diminished health in family caregivers. Previous research has found significant associations between family caregiving distress and declines in emotional, psychosocial and physical functioning [1-3]. One of the most highly investigated psychological conditions among the family caregiver population is depression. The large and growing body of research in this area has shown elevated depression scores in numerous samples of caregivers compared with their non-caregiver counterparts. In addition, several studies have found group differences across ethnicities on stress reactions, appraisals of the caregiving role, overall physical health, and level of depression [4-6]. Specifically, in comparison studies of African-American and non-Hispanic White caregivers, researchers have reported better cognitive coping strategies, more positive appraisal of caregiving role, and poorer overall physical health for African-American caregivers than their non-Hispanic White counterparts [2,4, 5,7,8]. Note, however, that this pattern has been inconsistent with regard to caregiver depression [9-12]. Pinquart and Sorensen [7] have argued

that discrepancies in sampling and measurement methods across caregiver studies have rendered differences in the severity of depression among ethnic groups difficult to interpret.

Studies in the general adult population have found differences in depressive symptom endorsement between African-Americans and non-Hispanic Whites, with African-Americans predominantly endorsing somatic symptoms such as weight changes or body aches and non-Hispanic Whites predominantly endorsing emotional depressive symptoms such as pessimism, self-blame, or suicidal ideation [13-16]. However, the empirical research examining this phenomenon is limited and does not address how and under what conditions such differential endorsement patterns occur. This lack of exploration into symptom interpretation precludes meaningful interpretation of depression scores on standardized instruments within and between ethnic groups.

Several "gold standard" instruments have been used to measure depression and depression treatment response over time. These include the Beck Depression Inventory [17,18], the Center for Epidemiological Studies Depression Scale [19], the Hamilton Rating Scale for Depression [20,21], and

the Zung Depression Rating Scale [22]. However, most of these measures are time-consuming, require a skilled mental health professional to administer, or were developed prior to the latest DSM revision and thus, do not measure aspects of criteria for depression found in the DSM-V or ICD-9 classifications.

More recently, both in clinical practice and research, the PHQ-9 has been used to assess depression in primary care settings and in persons with co-morbid physical health conditions. The PHQ-9, a self-report instrument, has 9 items that are based on the DSM-V diagnostic criteria for depression [23]. Each of the 9 items can be scored from 0 (not at all) to 3 (nearly every day). The instrument's validity and reliability as a diagnostic measure, utility in assessing depression severity, and ability to detect change over time have been reported in several studies [24,25]. Although the original validation study appeared to have a representative sample of African-American participants, the overall mean age of the participants (not specifically identified by race in the study results) was disproportionately younger than those typically found in the dementia caregiving population [26]. Thus, it is unclear whether the cutoff score for depression (PHQ-9 score > 9) is generalizable to the population of older dementia caregivers or, even more specifically, to older African-American adult caregivers. Limited analyses have been performed to better understand how individuals rate severity and assign meaning to items on the PHQ-9. In a previous study investigating differential item functioning (DIF; also known as differences in item endorsement) and mean depression scores on the PHQ-9, researchers reported no significant differences in mean PHQ-9 scores or item endorsement between African-Americans and non-Hispanic Whites [27,28]. Here again, this study's sample was significantly younger in age than the typical dementia caregiver and failed to address interpretive meaning of items across the two ethnic groups.

Although the aforementioned studies have underscored the importance of examining depression across races and ethnicities, the validity of their findings has been limited by the use of disparate measures of depression, perceived cultural uniformity of item interpretation, and wide variation in age and education of participants, as well as the language of test administration [29,30]. Failure to systematically test for differences on these factors may lead to erroneous conclusions about the nature and severity of depressive syndromes and inaccurate estimates of the prevalence of depression between and within target populations [30].

There are two primary reasons that item endorsement differences are found between racial and ethnic groups. The first is related to how one group may interpret an item's wording differently from another group. This typically results when an item's wording has different cultural relevance across ethnic/racial groups, or if an instrument has been translated inadequately. The second reason is related to differences in symptom expressions of a disorder across racial and

ethnic group[16]. For example and as noted earlier, previous research has suggested that when screened for depression, African Americans are more likely to endorse somatic symptoms as compared to non-Hispanic Whites, who tend to endorse more emotional symptoms [13,15]. Previous research also has shown that item complexity can affect participants' ratings on psychological assessment instruments. For example, the use of double- or triple-barreled items in depression inventories is especially problematic, rendering the overall meaning of such items difficult to interpret and in turn, potentially lowering the reliability and validity of results. In the case of the PHQ-9, this instrumentation artifact may lead participants to endorse responses that do not correspond to their actual feelings. It may also lead participants to ignore or fail to respond to an item if they feel that no valid response alternative is provided [31].

The primary objectives of this pilot study were to: (a) conduct a preliminary evaluation of differences in depression symptom interpretation (using the PHQ-9 items) across two samples of dementia caregivers, an African-American group and a non-Hispanic White comparison group; (b) perform item analyses examining differential symptom endorsement patterns (e.g., somatic versus cognitive) and severity ratings on items of the PHQ-9 within and between the two samples; and (c) assess the impact of item complexity on African-American and non-Hispanic White caregivers' item severity ratings.

## Method

### Participants

The present study included 14 participants (8 African-Americans and 6 non-Hispanic Whites), who were drawn from a grant-funded project evaluating the effects of cognitive-behavioral therapy on changes in depression and health status in African-American and non-Hispanic White caregivers [32,33]. Caregivers enrolled in the study met the following inclusionary criteria: (1) providing care for a family relative or significant other who met criteria for progressive dementia as specified by the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association [34]; (2) caring for a relative with progressive dementia who is 60 years of age or older; (3) being the primary care provider for the care recipient (CR); (4) spending a minimum of 6 hours per week in the provision of direct care to the person with progressive dementia; (5) scoring a minimum of 10 on the PHQ-9 [23]; and (6) being 18 years of age or older.

Caregivers were excluded from the study if they (1) endorsed fewer than 3 problems on the depression and disruptive behaviors factors of the modified version of the Revised Memory and Behavior Problem Checklist [35,36], (2) met criteria for psychotic disorder on the Mini-International Neuropsychiatric Interview [37] (3) met criteria for moderate or high suicide

risk on the M.I.N.I 5.0.0 (4) provided care for a person with dementia with a terminal diagnosis (six months as defined by hospice care); or (5) had a terminal diagnosis themselves.

The sample consisted of 1 male and 13 female caregivers with an average age of 65.14 (SD= 15.5) and an average education level of 14.14 years (SD = 4.7). Of the caregivers, 93% were married and on average had been providing care for their care recipient for 36.14 months (SD = 26.2)

### Primary Dependent Measure: PHQ-9

The PHQ-9 was developed to screen for depression in primary care settings [23]. It scores each of the nine DSM-V symptoms for depression on a scale of 0 (not at all) to 3 (nearly every day) over a two-week interval. A PHQ-9 score  $\geq 10$  has been found to have 88% sensitivity and 88% specificity for a diagnosis of major depressive disorder. PHQ-9 scores of 5, 10, 15, and 20 represent mild, moderate, moderately severe, and severe depression, respectively [24].

The nine global items on the PHQ-9 are introduced with the common stem, "over the last two weeks, how often have you been bothered by any of the following problems?" The symptoms queried (from here forward referred to as global items) are as follows:

1) Little interest or pleasure in doing things; 2) Feeling down, depressed, or hopeless; 3) Trouble falling or staying asleep, or sleeping too much; 4) Feeling tired or having little energy; 5) Poor appetite or overeating; 6) Feeling bad about yourself—or that you are a failure or have let yourself or your family down; 7) Trouble concentrating on things, such as reading the newspaper or watching television; 8) Moving or speaking so slowly that other people could have noticed, or the opposite being so fidgety or restless that you have been moving around a lot more than usual; and 9) Thoughts that you would be better off dead, or of hurting yourself in some way.

### Procedures

All caregivers were recruited from memory disorder clinics in Northern and Central Florida. Participants were informed of the purpose of the study and consented to participate in accordance with the research protocols approved by Florida State University Institutional Review Board. The current study was divided into two phases: initial screening for depression using the PHQ-9 (Phase I), and a repeat depression screening and semi-structured interview (Phase II). In Phase I, caregivers were "screened in" or "screened out" based on depression scores on the PHQ-9. Within two weeks of the initial screening, one of the authors contacted the caregivers over the telephone to complete a second structured interview in which the PHQ-9 was re-administered. The second interview followed the typ-

ical format and ordering of PHQ-9 items, but also used a cognitive interviewing approach [38] to collect additional information about participant responses. Cognitive interviewing is used to investigate how a targeted group understands, mentally processes, and responds to the materials presented (e.g., PHQ-9). Thus, during the second administration of the PHQ-9: (1) participants were asked about the personal meanings of each item (e.g. Do the words down, depressed, or hopeless have the same meaning to you? What does down/depressed/hopeless mean to you?); (2) participants made severity ratings of each item stem (e.g., Now having explored your meaning of down, depressed, and hopeless, how often over the last few weeks have you experienced them individually? Which would you rate as being worst: feeling down, feeling depressed, or feeling hopeless?); (3) participants were asked to provide illustrative examples of each item stem (e.g., what might cause you to feel down/depressed/hopeless? What might you look like when you are down/depressed/hopeless?); and (4) Participants were asked about perceived causes (e.g., what might cause you to feel down/depressed/hopeless?). This further inquiry allowed participants and the interviewer to explore the meaning and rationale for the severity rating of each item.

### Data Analysis

Semi-structured interviews were analyzed using the qualitative method techniques described by Huberman and Miles [39] to identify content themes by question and caregiver group. NVivo-8 qualitative analysis software was used for this purpose. The first author and two graduate assistants transcribed the digital interview recordings verbatim. After each interview was transcribed and entered into the NVivo-8 software, two coders (author and graduate assistant) analyzed the transcripts independently using the constant comparative method, open coding, and notating comments for each code [39]. The primary units of analysis were the African-American and non-Hispanic White caregiver's verbal responses to each question. The first author and the graduate assistant then met and discussed each code until they reached consensus. Following the coding process, the first author and graduate assistant analyzed the manuscripts for latent themes by reading one manuscript simultaneously, asking questions about the coded data, and identifying themes as they emerged. The first author then examined the dimensions of each theme. Themes occurring across all cases (i.e., occurring in each case) were identified as major themes [39].

While coding the team found that there were no real qualitative differences in six of the nine items (items 1, 3, 4, 7-9), but on the remaining three (items 2, 5, & 6) the qualitative differences were unambiguous, the team explored if there were any quantitative differences in the way the two groups responded on these items. We used Mann-Whitney U tests to ascertain if there were quantitative differences in the responses between the two groups on the three remaining items that we identified

as having qualitative differences in meaning and rating. The Mann-Whitney U is a robust, non-parametric test that calculates the differences in distributions between two independent samples. Mann-Whitney U's null hypothesis is that there is no difference between group distributions. Our data met the four assumptions of the Mann-Whitney U test (ordinal or continuous dependent variable, independent group/two category independent variable, independence of observations, and the two variables were not normally distributed). A significant Mann-Whitney U coefficient would be indicative of group differences on the 3 specific PHQ-9 global items (e.g. feeling down, depressed, or hopeless, poor appetite or overeating, and feeling bad about yourself...).

## Results

### Summary of Differences between Groups

#### Global Item 2

Global item 2 of the PHQ-9 consists of three different item stems (down, depressed, and hopeless) and in administering the PHQ-9 initially, all of the non-Hispanic Whites endorsed this item as at least "more than half of the days" whereas only half of the African Americans participants endorsed this item and for those who did endorse the item all rated it as "several days." In a structured interview re-administration format, the global item was broken down into its stemmed parts and the caregivers were asked to re-rate each stem separately. During this re-administration, African-American dementia caregivers tended to endorse feeling down "nearly every day" and to endorse feeling depressed or hopeless "not at all." In contrast, non-Hispanic White caregivers endorsed feeling down and feeling depressed "nearly every day," and feeling hopeless "several days."

**Major themes for global item 2: Differences between groups. "Lacking energy and can't escape" vs. "Sadness."** The theme "lacking energy" emerged as the meaning of feeling down within the African-American caregiver group. African-American caregivers tended to describe feeling down as "being tired (from a 58 year-old female caregiver)," "sinking (from another 58 year-old female caregiver)," "can't get up (from a 62 year-old female caregiver)," "real tired (from a 65 year-old caregiver)," "don't have energy (from a 44 year-old female caregiver)," "giving up at that time, not able to get it going (from a 48 year-old female caregiver)," and "can't get up but an opportunity to get better (from a 57 year-old female caregiver)."

The theme "can't escape" emerged for the meaning of feeling depressed among the African-American caregivers. Most African-American caregivers tended to describe depression as being locked in a situation from which they could not extricate themselves. For example, a 58 year-old caregiver stated, "De-

pressed to me is a mental state in which one feels trapped." Likewise, a 62 year-old caregiver said, "When I feel like I am going further and further down and just cannot get out of prison," and a 65 year-old caregiver said, "not being able to feel, no matter how hard you try." Other examples of this come from a 59 year-old caregiver who felt that feeling depressed meant, "Everything is a hassle, nothing can be done," from a 44 year-old caregiver who said, "I feel like I can't change it [the situation] physically or mentally," and from a 48 year-old caregiver who said, "No energy to figure it out." A 57 year-old caregiver similarly noted about feeling depressed: "In a state, it may be temporary, but it feels permanent."

In contrast to the themes of "lacking energy" and "can't escape" that emerged for the African-American caregivers, the theme of "sadness" emerged as a description of down and depressed for non-Hispanic White caregivers. When exploring the overall meaning of the item stems of global item 2 (down, depressed, and hopeless), the non-Hispanic White caregivers indicated meanings such as: "feeling glum (from a 52 year-old caregiver)," "sad (from a 53 year-old caregiver)," "unhappy (from a 54 year-old caregiver)," and "sad with a heaviness (from a 55 year-old caregiver)."

**"Giving up."** For African-American dementia caregivers (n = 8), all except one answered that they never felt "hopeless." Based on discussions with these participants, the caregivers attributed their absence of hopelessness to their faith in God. When asking them to explore the meaning of the word "hopeless," African-American caregivers indicated meanings such as, "it is the dredges, being on the fringe of suicide...;" "complete devastation;" "giving up on life itself;" "nothing can be done about it: void;" "I give up;" and "absolutely no way."

No consistent, overarching theme emerged for the non-Hispanic White caregiver group for the term "hopeless." However, we found some similarities between responses of the non-Hispanic White caregiver sample and responses of the African-American caregiver sample. Most notably, 50% of the non-Hispanic White caregivers also described the term 'hopeless,' as "giving up." For example, a 54 year-old non-Hispanic White caregiver defined hopeless as, "not being able to see beyond...no reason to go on." A 55 year-old non-Hispanic White caregiver thought there was, "no solution;" "no light at the end of the tunnel." Finally, a 51 year-old caregiver described hopeless as, "don't believe in God, no faith in a supreme being." This half of the non-Hispanic White sample viewed hopelessness similarly to the African-American caregivers.

However, the other half of the non-Hispanic White sample seemed to describe hopeless as a form of "rejection or pain." Some examples of these non-Hispanic White caregiver descriptions of hopeless include, "completely rejected," "feeling unloved," and "complete agony."

**Differences in assigned meaning and impact of item complexity for global item 2.** When item 2 was divided into its three item stems (“how often have you felt down,” “how often have you felt depressed,” and “how often have you felt hopeless”), all non-Hispanic White and African-American caregivers positively endorsed feeling down more often than depressed or hopeless.” Only one of eight (12.5%) African-American caregivers reported feeling hopeless several days a week, whereas five of six (83%) non-Hispanic White caregivers reported feeling hopeless, with two endorsing feeling hopeless on several days during the past two weeks and the others endorsing experiencing the item “nearly every day.”

When asked to describe what down, depressed and hopeless mean, African-American caregivers tended to describe “down” related to their energy level and “depressed” as an inability to escape. For the non-Hispanic White caregiver group, the terms down and depressed produced the item theme: ‘sadness.’ When we explored the meaning of “hopeless” with the two groups, the African-American caregiver group tended to describe this as “giving up.” Two themes emerged for the term, hopeless, among the non-Hispanic White caregiver group, largely because the group was split in their description of the term. Due to this split in description we began to explore the responses of the groups more carefully.

During this deeper exploratory examination, we found similarities and distinct differences between the groups. Several of the non-Hispanic White group members, similar to their African-American counterparts, described hopeless as ‘giving up.’ Interestingly, the other half of the group described hopeless as ‘rejection or pain.’ We further explored whether their descriptions were related to their response patterns and found that those who described hopeless as ‘giving up,’ regardless of their group membership, tended not to endorse the item stem ‘feeling hopeless.’ In all cases, this was attributed to the caregivers’ belief in God. On the other hand, the non-Hispanic White caregivers who described hopeless as ‘rejection or pain’ tended to positively endorse feeling hopeless and directly attributed this feeling either to their loved one’s “medical condition” or due to the “unrelenting course of the disease.”

Furthermore, in review of item response difference, a Mann Whitney U analysis identified a significant difference in the responses between African-American and non-Hispanic White caregivers on item 2 (feeling down, depressed, or hopeless). Non-Hispanic Whites demonstrated higher scores than African Americans,  $U = 2.03.5$  ( $Z = -2.507$ ),  $p = .012$  and the difference between non-Hispanic Whites and African Americans was moderate to large ( $r = .67$ ). We used the Holm-Bonferroni correction for multiple comparisons; alpha/m, [40,41] and the alpha value associated with this score maintained significance.

## Global Item 5

Global item 5 of a PHQ-9 consists of two item stems (poor appetite and overeating). The qualitative responses of the participants indicated some interesting qualitative differences. In the initial administration of the PHQ-9, 75% of the African-American caregivers tended to endorse the item as “not at all”. In the re-administration of the instrument, African Americans endorsed “overeating” as “more than half the days” and poor appetite as “not at all” (75%;  $n=8$ ). One African-American caregiver endorsed poor appetite, but he indicated that his loss of appetite was related to alcoholic consumption patterns that he used primarily to cope with his caregiving situation.

Conversely, non-Hispanic White caregivers overwhelmingly endorsed the item in the first administration (100%). In the re-administration of the instrument, non-Hispanic Whites endorsed poor appetite (83%;  $n=6$ ) “nearly every day” and endorsed overeating as “not at all” (100%;  $n=6$ ). Overeating was not endorsed by any of the non-Hispanic White caregivers, whereas overeating was endorsed by almost all of the African-American caregiver participants.

**Major Themes for Global Item 5: Differences between groups.** As with global item 2, caregivers were asked to explore the meaning of the individual item stems for global PHQ-9 item 5 (poor appetite and overeating).

**“No desire” vs. “No energy.”** The theme, “no desire” materialized as a meaning of poor appetite within the African-American caregiver group. African-American caregivers described poor appetite as “not wanting anything to eat” (African-American caregiver-female, age 58); “not wanting to eat correctly” (African-American caregiver-female, age 62); “not wanting any food” (African-American caregiver-male, age 65); and “don’t want anything...” (African-American caregiver-female, age 57).

For non-Hispanic White caregivers, the theme “no energy” emerged as a meaning for poor appetite. One non-Hispanic White caregiver indicated that poor appetite means “someone having a distaste for food because they are tired” (non-Hispanic White caregiver, age 53); “not eating because of no energy... lazy” (non-Hispanic White caregiver, age 54); “no energy” (non-Hispanic White caregiver, age 55); “not wanting to cook...exhaustion because you are responsible for everything” (non-Hispanic White caregiver, age 64); “not eating because you are tired or distracted” (non-Hispanic White caregiver, age 51); and “not eating because of physical and emotional stress... exhausted” (non-Hispanic White caregiver, age 52).

**“Pleasure seeking” vs. “Stress-reducing.”** The theme, “pleasure seeking,” emerged as one meaning of overeating among the African-American caregivers. Six of eight African-American caregivers tended to describe overeating as pleasing oneself because they enjoyed eating (i.e., pleasure seeking). For

example, a 58 year-old African-American caregiver said of the motivation to overeat, "eating in abundance because the food is good." Other African-American caregivers described their experiences of overeating as, "gorging on food...my husband likes to eat, especially desserts;" "when someone eats way too much because they like what they are eating;" "eating even after you have reached satisfaction;" "Everything I cook tastes good that is often why I keep eating;" "eating beyond being full usually because it is good;" and "eating an overdose of food... it's good."

Comparatively, the theme, "stress-reducing," emerged for those non-Hispanic White caregivers who endorsed overeating, as all of them indicated that overeating is related to some sort of stress. A 53 year-old non-Hispanic White caregiver said of the link between eating and stress, "eating beyond a point of satisfaction... typically because you are an emotional wreck." Other non-Hispanic White caregivers made similar statements about stress and eating, including a 54 year-old who answered, "gorging myself to relieve stress," a 64 year-old who said, "eating junk and snack after meals for comfort," a 51 year-old who answered, "eating too much due to stress related issues," and a 52 year-old who said, "always feeling hungry and never satisfied because you need relief."

**Differences in assigned meaning and impact of item complexity for global item 5.** A review of responses to Item 5 suggests that both non-Hispanic White and African-American caregivers had a tendency to express the stress of caregiving through dietary changes. However, when separating the parts of the mutually exclusive components of the item, the non-Hispanic White group was more likely to experience poor appetite, whereas the African-American caregiver group tended to experience overeating.

When asked to describe their meaning associated with "poor appetite or overeating" non-Hispanic White caregivers described overeating in terms of stress reduction, whereas African-American caregivers tended to use descriptors related to their physical desire. Interestingly, while all of the non-Hispanic White caregivers linked changes in eating habits to emotional distress, only one African-American caregiver even mentioned stress when probed about eating patterns.

There was also significant difference in the response patterns on item 5 (Poor appetite or Overeating) between African-American and non-Hispanic White caregivers, with non-Hispanic Whites scoring lower than African Americans,  $U = 1.0$  ( $Z=2.803$ ,  $p = .005$ ) and the difference between non-Hispanic Whites and African Americans was large ( $r = .75$ ). Again, even after corrections for multiple comparisons this item remained significant [40,41].

## Global Item 6

Global item 6, consists of four item stems: "Feeling bad about yourself," "feeling that you are a failure," and "feel that you have let yourself or feeling you have let your family down." In the initial administration of the PHQ-9, 37.5% ( $n = 3$ ) of the African-American caregiver sample endorsed the item, whereas almost all (83%,  $n = 5$ ) of the non-Hispanic White caregivers endorsed experiencing the global item at least several days a week since they began caregiving. In the PHQ-9 re-administration in a structured interview format, only three of the African-American caregivers endorsed any of the item stems. The only item stem endorsed by these African-American caregivers was "feeling like you let your family down." In comparison, the non-Hispanic White caregivers all endorsed one of two item stems as 'several days a week' during the PHQ-9 re-administration: either feeling bad about yourself (67.7%;  $ns = 4$ ) or letting yourself down (33.3%;  $ns = 2$ ). No member in the non-Hispanic White group endorsed "feeling like you let your family down."

**Major themes for global item 6: Differences between groups.** When the dementia caregivers were asked to explore the meaning of the individual item stems of PHQ-9 item 6 (Feeling bad about yourself, feeling that you are a failure, feeling like you have let yourself down, or feeling like you have let your family down), we were able to identify several interesting thematic patterns.

**"External self-disappointment" vs. "Internal self-disappointment."** In both the (African-American) and the non-Hispanic White caregiver groups, the themes of 'external disappointment' and 'internal disappointment' emerged for responses to the item stems "feeling bad about yourself," and "letting yourself down." There were several similarities in the descriptors used across the groups representing these themes. There were also some interesting differences in responses, particularly regarding the theme of external self-disappointment.

The African-American caregivers' descriptions of 'external disappointment' were as follows: "not doing something I should have;" "struggling with our finances;" "not fulfilling my obligations;" "doing nothing;" "not completing what was planned;" and "not doing my best." In contrast, non-Hispanic White caregiver descriptors of 'external self-disappointment' included such things as: "seeing wrinkles all over my body;" "looking old;" "aging;" "external beauty fading;" and "not being able to make a success out of what you wish for."

In these descriptors of 'external self-disappointment,' we observed that the African American participants' statements related to the lack of performance or success. However, within the non-Hispanic White participants we find 'external self-disappointment' in reference to feeling bad about self that is pure-

ly based on physical features (e.g. “seeing wrinkles all over my body;” “looking old;” “aging;” and “external beauty fading”).

Examples of the ‘internal self-disappointment’ descriptors within the African-American caregiver group were statements like: “low self-esteem;” “self-disapproval;” and “disappointed in myself.” Non-Hispanic White caregiver descriptors of this same theme included remarks like: “being disappointed in my conduct;” “being to blame for what is going wrong;” “feeling you are to blame or that it is your fault;” “things not ending up the way I envisioned;” “extreme lack of performance;” “an inability to turn things into what you wish;” “giving up;” “my life not going the way I thought it would;” and “disappointment.”

Limited descriptions were provided by both the African-American and the non-Hispanic White groups with regard to the item “feeling like you have let your family down.” African-American caregivers made statements such as, “my family being disappointed in my performance;” “trying to make someone feel better but you can’t;” and “knowing that I should have done better by my loved ones.” Only one descriptor was provided by the non-Hispanic White group for this item. This was, “not measuring up to my family’s expectations of my performance.” Based on this limited description of the stem by the non-Hispanic White group, no theme could be generated.

**Differences in assigned meaning and impact of item complexity for global item 6.** Item 6, “feeling bad about yourself or that you are a failure, or feeling that you have let yourself or your family down,” was positively endorsed by approximately half of the African-American caregiver sample, whereas almost all of the non-Hispanic White caregivers positively endorsed this item. During the structured interview and separating out the item stems, 75% ( $n = 6$ ) of the African-American caregivers did not endorse any of the item stems, and of the two who did, each solely endorsed “feeling like you let your family down.” Conversely, almost the entire non-Hispanic White caregiver group endorsed “feeling bad about yourself,” and one endorsed “letting yourself down.” None of the non-Hispanic White group members endorsed “feeling like you let your family down.” When asked to state in their own words what these item stems meant, non-Hispanic White caregivers again reported internalizing language in their definitions of the words, whereas African-American caregivers tended to use externalizing language. This may further highlight differences in how this symptom might be expressed between the groups.

We also found differences in responses between the two groups in global item 6. The Mann-Whitney U showed that non-Hispanic Whites scored higher on global item six (feeling bad about yourself...) than did African Americans,  $U = 4.5$  ( $Z=1.979$ ),  $p = .048$ ) and the difference between non-Hispanic Whites and African Americans was moderate ( $r = .53$ ); this difference remained significant after corrections for multiple

comparisons.

### Limitations

As with any study using a small sample and participants from one particular region of the country, there were limitations regarding the generalizability of the findings. The sample primarily included residents of Northern and Central Florida, and thus might not be representative of African-American and non-Hispanic White dementia caregivers in other regions of the U.S. Additionally, the use of non-probability sampling makes it possible that the groups of caregivers differed on important characteristics other than their racial and ethnic heritage. Therefore, the findings of this study must be interpreted with caution. Another limitation is that participants reported and rated their symptoms of depression over the telephone. Inherent to this format is that interviewers could not incorporate participant nonverbal responses about their depressive symptoms, particularly affect and/or other body language indicators often used in the assessment of depressive symptoms.

### Discussion and Implications

The purpose of this study was to explore qualitative differences in item meaning on the PHQ-9 between African-American and non-Hispanic White caregivers, as well as to ascertain item complexity effects on rating strategies, and to explore if there were item response differences associated with qualitative meaning. Despite the associated limitations outlined in the previous section, the findings suggest potential cultural patterns in depressive symptoms and PHQ-9 responses that warrant further investigation of their generalizability and impact on clinical practice with diverse caregivers. In the three items that demonstrated qualitative differences in meaning, we also demonstrated quantitative response differences as well as changes in scoring when singularly submitting items stems to participants. However, the fact that 6 of the nine items demonstrated no significant qualitative differences, could possibly mean that there is error in our findings or that these results could have happened by chance, and given our sample size this is a realistic possibility. Yet, one could also argue the opposite, that our sample size just was not large enough to capture the true differences that are found in the other six items, and given the endorsement changes when items were separated out, coupled with the quantitative differences, and associated strength of the relationships even after corrections were made; this position also has support. Lastly, one could argue that there truly are no differences between the two groups on item meaning of the remaining items. No matter which position you take, these findings do warrant further investigation in larger sample sizes. If these findings, (item meaning differences, item endorsement changes, and item response differences), were to be replicated in a larger sample, such would

have implications for the future use of the PHQ-9 for screening and assessment as well as the potential need to restructure the instrument, particularly for continued implementation with diverse groups. If reconstructing were warranted, the PHQ-9 could be restructured similar to how the authors in this study broke a part the global items down into the item stems and presented them individually to the participants. For example, global item 2 in the PHQ-9, "Feeling down, depressed, or hopeless" would become 3 global items: 1. Feeling down; 2. Feeling depressed; and 3. Feeling hopeless. This will allow the respondent to rate symptoms individually and minimize item complexity. We found that this may be important particularly due to endorsement patterns of certain groups (e.g. African Americans not ascribing hopelessness and non-Hispanic Whites not ascribing overeating). Since, as practitioners we are still only interested in identifying if the 9 items associated with the DSM-5 are measured, the developers could instruct the scorer to recombine previously separated items (Feeling down, depressed, hopeless) and assign the highest rating of the 3 items that all measure sad mood. For example, if the respondent rated down as a 3, depressed as a 1, and hopeless as a 0, the scorer would give the respondent a 3 for sad mood.

The differences in item responses observed in this study also appear to be related to assigned meaning. Major themes for the sample were the differential roles of internal and external concerns, as well as the roles of stress and pleasure in coping behaviors and depressive symptoms for African-American and non-Hispanic White caregivers. Again, future research should explore these themes with larger samples of caregivers as well as differences that may be found in racial/ethnic groups from different geographical regions; as well as explore measurement invariance in depression screeners among diverse groups of dementia caregivers.

Some might wonder what difference it makes if people differentially endorse symptoms or intend to only endorse one aspect of the item (e.g. overeating rather than poor appetite). The problem lies in meaning and decision-making approaches. If combining "down, depressed, and hopelessness" together in one item causes individuals not to endorse the item because they would never say that they are hopeless, but when you separate out the stems of the item they endorse feeling down and depressed more than half the days, then there is a fundamental problem with the item and the overall measurement of the construct for these individuals and potentially the group. For example, in our study African-American caregivers were less likely than their non-Hispanic White counterparts to endorse symptoms of feeling down, depressed, or hopeless when these items were assessed together. However, when these item stems were separated into individual items, the African-American caregiver group endorsed "feeling down" much more often and rated the item more negatively. Thus, it would appear that coupling the items "feeling down, depressed, or hope-

less" in the same global item impacted the likelihood of African-American caregivers' endorsement of the item and rating the item as severe. Additionally, despite the fact that non-Hispanic White caregivers tended to endorse this same item more often than their African-American counterparts, we found that when the item stems were separated, non-Hispanic White caregivers were also more likely to endorse feelings of down or depressed over hopelessness, and when they did express feelings of hopelessness, it was exclusively related to a reaction to their loved ones medical condition and disease burden.

Focusing on the overall score and minimum endorsement of the first two items (little interest and feeling depressed) as a requirement for a diagnosis of major depression becomes problematic when endorsement of those items are strongly influenced by the interpretation of the item stems and the complexity of the items. In this study, we found that there are differences in item endorsement on the PHQ-9 among different dementia caregiver groups when the items are administered globally compared to when they are administered by individual item stems. If the caregiver is less likely to endorse feeling down if it is coupled with feeling hopeless then it is likely that at least one of the main criteria for depression will not be met and recognition of depression in the caregiver could be missed. Whether this endorsement pattern is due to socio-cultural factors could not be determined based on this study, but it is clear that hopelessness was not readily endorsed by African-American caregivers and that this item coupled with down or depressed decreased the likelihood that the item would be endorsed and rated as severe.

There were also differences in which symptoms each group identified as more detrimental to their everyday functioning. For one group having even low levels of overeating was more functionally disabling than having high levels of mood disturbances. This finding draws attention to diagnostic criteria and differential application across groups. Again, for an individual to warrant treatment at least one of two symptoms must be positively endorsed at least half of the days – anhedonia and/or sadness. This criterion assumes that all groups are functionally impacted the same by the presence of certain symptoms. However, this study's findings suggest potential criterion differences across groups. For instance, although both the African-American and non-Hispanic White caregivers agreed that experiencing overeating would be more detrimental to their functioning, only African-American caregivers positively endorsed overeating as a symptom. This may suggest the presence of this symptom as a unique criterion indicator for depression in this group. Further research is needed to examine overeating as a symptom of depression among African-American caregivers, as the small sample size of this study allows for the possibility that this finding is an artifact of a shared trait among the African-American caregivers that is not culture-specific, such as weight (weight not being culture bound

and may just be coincidental).

Additionally and as mentioned in the literature review, previous research has shown that populations of color are more likely to positively endorse somatic symptoms of depression such as overeating over affective symptoms of depression such as depressed mood or hopelessness [13-16]. Our study's findings seem to support these findings but seem to add some meaning to the differences found. For instance, African American caregivers were less likely to endorse feeling down when it was in a combined item stem with feeling hopeless. This might add some explanation as to why African Americans are less likely to endorse affective symptoms of depression as a product of item complexity and ascribed meaning.

## Conclusion

Concern for the cultural appropriateness and sensitivity of instruments and practices is a trademark feature of social work research and practice. It is imperative that measurement instruments address the complexity of culture in their development. Depression among dementia caregivers is a great burden on their functioning which can impair their caregiving abilities and overall health. For some groups, particularly African Americans, caregiving is almost exclusively provided by the loved ones of the dementia patient and they are less likely to have resources available to aid them [42-44]. Findings of previous studies highlight the need to assess depression among dementia caregivers. The current study's exploratory results provide some indication that non-Hispanic White and African-American caregivers of family members with dementia express depressive symptomatology in different ways. When assessing depressive symptoms in dementia caregiving groups, it is important to note the possibility of meaning interpretation and ratings differences not only between groups, but also based on item complexity. For example, several items in the PHQ-9 contain stems of opposite meaning, such as poor appetite and overeating. If one participant had a poor appetite nearly every day, yet never overate, they could possibly average their score seeing that the two options are opposite and rate their global score as more than half the days rather than rating it as nearly every day. Thus, this type of item complexity has the potential to cause considerable variation both within and between caregiving groups on their use of strategies in arriving at a global item rating.

## References

1. Aranda MP. Racial and ethnic factors in dementia care-giving research in the US. *Aging and Mental Health*. 2001, 5(suppl. 1): S116-S123.
2. Pinquart M, Sörensen S. Associations of stressors and uplifts of caregiving with caregiver burden and depressive mood: A meta-analysis. *J Gerontol B Psychol Sci Soc Sci*. 2003, 58B(2): P112-P128.
3. Roth DL, Haley WE, Owen JE, Clay OJ, Goode KT. Latent growth models of the longitudinal effects of dementia caregiving: a comparison of African American and White family caregivers. *Psychol Aging*. 2001, 16(3):427-436.
4. Aranda MP, Knight BG. The influence of ethnicity and culture of the caregiver stress and coping process: a sociocultural review and analysis. *Gerontologist*.1997, 37(3): 342-354.
5. Connell CM, Gibson GD. Racial, ethnic, and cultural differences in dementia caregiving: review and analysis. *The Gerontologist*.1997, 37(3): 355-364.
6. Lampley-Dallas VT. Research issues for minority dementia patients and their caregivers: what are the gaps in our knowledge base? *Alzheimer Dis Assoc Disord*.2002, 16(Suppl. 2): S46-S49.
7. Pinquart M, Sörensen S. Ethnic differences in stressors, resources, and psychological outcomes of family caregiving: a meta-analysis. *Gerontologist*.2005, 45(1): 90-106.
8. Calderon V, Tennstedt SL. Ethnic differences in the expression of caregiver burden: Results of a qualitative study. *Journal of Gerontological Social Work*.1998, 30(1-2): 159-178.
9. Knight BG, Silverstein M, McCallum TJ, Fox LS. A sociocultural stress and coping model for mental health outcomes among African American caregivers in Southern California. *Journals of Gerontology Series B*.2000, 55(3): 142-150.
10. Knight BG, McCallum TJ. Heart rate reactivity and depression in African- American and white dementia caregivers: reporting bias or positive coping? *Aging and Mental Health*. 1998, 2(3): 212-221.
11. Haley WE, West CA, Wadley VG, Ford GR, White FA et al. Psychological, social, and health impact of caregiving: a comparison of Black and Caucasian dementia family caregivers and noncaregivers. *Psychology and Aging*.1995, 10(4): 540- 552.
12. Haley WE, Roth DL, Coleton MI, Ford GR, West CA, et al. Appraisal, coping, and social support as mediators of well-being in Black and White family caregivers of patients with Alzheimer's disease. *J Consult Clin Psychol*.1996, 64(1):121-129.
13. Ayalon L, Young M A. Comparison of depressive symptoms in African Americans and Caucasian Americans. *Journal of Cross-Cultural Psychology*.2003, 34(1): 111-124.
14. Adebimpe VR. Overview: White norms and psychiatric di-

- agnosis of Black patients. *American Journal of Psychiatry*.1981, 138(3): 279-285.
15. Brown C, Schulberg HC, Madonia MJ. Clinical presentations of major depression by African Americans and Whites in primary medical care practice. *J Affect Disord*.1996, 41(3): 181-191.
16. Kim G, Chiriboga DA, Jang Y. Cultural equivalence in depressive symptoms in White, Black, and Mexican-American Adults. *J Am Geriatr Soc*.2009, 57(5): 790-796.
17. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatry*.1961, 4: 561-571.
18. Beck AT, Steer R A, Brown G K. Beck Depression Inventory manual (2nd ed.). San Antonio, TX: Psychological Corporation, 1996.
19. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement*.1977, 1(3): 385-401.
20. Hamilton M. A rating scale for depression. *Journal of Neurology, Neurosurg Psychiatry*.1960, 23: 56-62.
21. Hamilton M. Development of a rating scale for primary depressive illness. *Br J Soc Clin Psychol*.1967, 6(4): 278-296.
22. Zung WWK. A self-rating depression scale. *Arch Gen Psychiatry*. 1965, 12: 63-70.
23. Kroenke K, Spitzer R. The PHQ-9: A new depression diagnostic and severity measure. *Annals of Psychiatry*.2002, 32:509-515.
24. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: Validity of a brief depression severity measure. *J Gen Intern Med*. 2001, 16(9): 606-613.
25. Löwe B, Gräfe K, Zipfel S, Witte S, Loecherer B et al. Diagnosing ICD-10 depressive episodes: Superior criterion validity of the Patient Health Questionnaire. *Psychother Psychosom*. 2004, 73(6): 386-390.
26. Spitzer RL, Kroenke K, Williams JB. Patient Health Questionnaire study group, validity and utility of a self-report version of PRIME-MD: The PHQ primary care study. *Journal of the American Medical Association*.1999, 282: 1737-1744.
27. Huang FY, Chung H, Kroenke K, Delucchi KL, Spitzer RL. Using the Patient Health Questionnaire-9 to measure depression among racially and ethnically diverse primary care patients. *J Gen Intern Med*.2006, 21(6): 547-552.
28. Löwe B, Unützer J, Callahan CM, Perkins AJ, Kroenke K. Monitoring depression treatment outcomes with the Patient Health Questionnaire-9. *Med Care*. 2004, 42(12): 1194-1201.
29. Hui HC, Triandis HC. Measurement in cross-cultural psychology: a review and comparison of strategies. *Journal of Cross-Cultural Psychology*.1985, 16(2): 131-152.
30. Jones RN. Identification of measurement differences between English and Spanish language versions of the Mini-Mental State Examination: Detecting differential item functioning using MIMIC modeling. *Med Care*. 2006, 44(Suppl. 3): S124-S133.
31. Clark LA, Watson D. Constructing validity: Basic issues in objective scale development. *Psychological Assessment*.1995, 7(3): 309-319.
32. Glueckauf RL, Davis WS, Allen K, Chipi P, Schettini G et al. Integrative cognitive-behavioral and spiritual counseling for rural dementia caregivers with depression. *Rehabil Psychol*. 2009, 54(4): 449-461.
33. Glueckauf RL, Davis WS, Willis F, Sharma D, Gustafson DJ. Telephone-based, cognitive-behavioral therapy for African-American dementia caregivers with depression: Initial findings. *Rehabil Psychol*. 2012, 57(2): 124-139.
34. McKhann G, Drachman D, Folstein M, Katzman R, Price D et al. Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA work group. *Neurology*, 34(7): 939-944.
35. Teri L, Truax P, Logsdon R, Uomoto J, Zarit S et al. Assessment of behavioral problems in dementia: The Revised Memory and Behavior Problems Checklist (RMBPC). *Psychol Aging*.1992, 7(4): 622-631.
36. Burgio L, Stevens A, Guy D, Roth DL, Haley WE. Impact of two psychosocial interventions on White and African American family caregivers of individuals with dementia. *Gerontologist*.2003, 43(4): 568-579.
37. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry*.1998, 59(Suppl. 20): 22-32.
38. Willis GB. Cognitive interviewing: A tool for improving questionnaire design. Thousand Oaks, CA: SAGE Publications, 2005.

39. Huberman A M, Miles MB. The qualitative researcher's companion. Thousand Oaks, 2002.
40. Aickin M, Gensler H. Adjusting for multiple testing when reporting research results: The Bonferroni vs Holm method. *Am J Public Health*.1996, 86(5): 726-728.
41. Holm SA. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*.1979, 6: 65-70.
42. Rabinowitz YG, Saenz EC, Thompson LW, Gallagher-Thompson D. Understanding caregiver health behaviors: depressive symptoms mediate caregiver self-efficacy and health behavior patterns. *Am J Alzheimers Dis Other Demen*. 2011, 26(4): 310-316.
43. Lovett SB, Gallagher D, Benedict A, Priddy JM, Poulsen SL. Relationships between coping style and caregiver stress. *Society of Behavioral Medicine 7th Annual Meeting*, San Francisco, CA, 1986.
44. Schulz R, Visintainer P, Williamson GM. Psychiatric and physical morbidity effects of caregiving. *J Gerontol*.1990, 45(5): 181-191.