Published in final edited form as:

Psychol Aging. 2003 December ; 18(4): 886–895. doi:10.1037/0882-7974.18.4.886.

The Revised Memory and Behavior Problems Checklist— Nursing Home: Instrument Development and Measurement of Burden Among Certified Nursing Assistants

Rebecca S. Allen,

Department of Psychology and Center for Mental Health and Aging, University of Alabama at Tuscaloosa

Louis D. Burgio,

Department of Psychology and Center for Mental Health and Aging, University of Alabama at Tuscaloosa

David L. Roth,

Department of Biostatistics, University of Alabama at Birmingham

Rachael Ragsdale,

Center for Mental Health and Aging, University of Alabama at Tuscaloosa

John Gerstle,

Center for Mental Health and Aging, University of Alabama at Tuscaloosa

Michelle S. Bourgeois,

Department of Communication Disorders, Florida State University

Katinka Dijkstra, and

Department of Psychology, Florida State University

Linda Teri

Department of Psychosocial and Community Health, University of Washington School of Nursing

Abstract

Confirmatory factor analysis of the Revised Memory and Behavior Problems Checklist—Nursing Home (RMBPC) replicated the factor structure of the community-based RMBPC (L. Teri et al., 1992). The reliability of the total score was high as indexed by estimates of internal consistency ($\alpha = .95$), test–retest reliability ($r = .86$), and interrater reliability between 2 interviewers ($r = .88$). Notably, the interrater reliability between 2 independent certified nursing assistants (CNAs) regarding residents' behavior problem frequency was more modest ($r = .46$), possibly reflecting the degree to which resident behaviors capture individual CNA's attention. This may have

Correspondence concerning this article should be addressed to Rebecca S. Allen or Louis D. Burgio, University of Alabama, Department of Psychology and The Center for Mental Health and Aging, Box 870348, Tuscaloosa, Alabama 35487-0348. rsallen@bama.ua.edu or lburgio@sw.ua.edu.

Portions of this article were presented at the 52nd annual scientific meeting of the Gerontological Society of America, San Francisco, California, November 1999.

implications for the interpretation of data from the Minimum Data Set. CNAs reported moderately severe burden associated with behavior problems in 47% of residents under their care.

Estimates of the prevalence of behavior problems among individuals with dementia residing either in nursing homes or in the community range from 42% (American Health Care Association, 1999) to 86% (Deutsch & Rovner, 1991; Ryden, Bossenmaier, & McLachlan, 1991). Such diverse estimates result from complications researchers have encountered in reaching consensus regarding precise definitions of behavior problems and variability in methods of measurement (Burgio, 1997; Davis, Buckwalter, & Burgio, 1997; Vance et al., 2003). One of the more prevalent “syndromes” of problematic behavior has been labeled *agitation* (Cohen-Mansfield & Billig, 1986), operationally defined as inappropriate verbal, vocal, or motor activity not solely resulting from the physical needs of the agitated individual. Researchers have proposed that agitation results from an incongruent person–environment fit (Algase et al., 1996; Beck & Vogelpohl, 1999; Lawton, 1989). That is, when biopsychosocial needs are not met, individuals with compromised internal resources (i.e., neurological and cognitive decline) may react to situational factors in the physical and social environment with behavioral agitation.

Accurate assessment of behavior problems is critically important because they have direct clinical implications for individuals with dementia and for their caregivers (Algase et al., 1996; Cohen-Mansfield, 1994). The most common means of assessing behavior problems in the nursing home is through paper-and-pencil surveys (Cohen-Mansfield, 1995). Surveys used in nursing homes have asked staff to report the frequency of behavior problems (Cohen-Mansfield Agitation Inventory [CMAI]; Cohen-Mansfield, Marx, & Rosenthal, 1989) or merely to report their occurrence/nonoccurrence (Memory and Behavior Problems Checklist—Nursing Home [MBPC–NH]; Wagner, Teri, & Orr-Rainey, 1995a, 1995b). These questionnaires have been administered either to certified nursing assistants (CNAs; Cohen-Mansfield et al., 1989) or to family members, registered nurses, and licensed practical nurses (Wagner et al., 1995a, 1995b). When using questionnaires, the best estimates of the frequency of behavior problems likely will be obtained from CNAs because those staff members have the most direct contact with residents (Burgio, Engel, Hawkins, McCormick, & Scheve, 1990; Chappell & Novak, 1994; Tellis-Nayak & Tellis-Nayak, 1989).

One of these measures, the MBPC–NH (cited in Wagner et al., 1995a, 1995b) was adapted from Zarit and Zarit's (1983) Memory and Behavior Problems Checklist (MBPC), designed for use among community caregivers. The MBPC was subsequently revised (Revised Memory and Behavior Problem Checklist [RMBPC]; Teri et al., 1992). The RMBPC has been used widely in various studies in the community because it not only provides data on the frequency of behavior problems, but also provides information on caregivers' perceptions of burden associated with those behaviors. Among community samples, burden has been defined as appraised distress in response to caregiver experiences (Lawton, Kleban, Moss, Rovine, & Glicksman, 1989; Vitaliano, Russo, Young, Teri, & Maiuro, 1991; Zarit, Reever, & Bach-Peterson, 1980). Behavior problems exhibited by the person with dementia have consistently been associated with perceived burden (Teri et al., 1992; Vitaliano et al., 1991).

Although burden appeared as a research construct 20 years ago (Zarit et al., 1980), researchers have not attempted to assess burden in nursing homes, possibly due to the difficulties encountered in generating precise definitions of the burden experienced by formal caregivers. However, we know that resident agitation has been associated with staff distress (Chappell & Novak, 1994; Everitt, Fields, Soumerai, & Avorn, 1991; J. E. Fisher, Fink, & Loomis, 1993) and avoidance of residents, which can impact the quality of resident care (Burgio, 1997; Cohen-Mansfield, 1995). Based on reports by staff, physical and verbal aggression, refusal to cooperate, restlessness, and verbal disruption (i.e., repetitive vocalizations) are the most difficult behavior problems to manage and can affect the physical health of staff, including days sick in bed, days of work missed due to illness, and use of pain relievers (Chappell & Novak, 1994).

The purpose of this study was to modify the MBPC–NH, previously used in nursing homes to measure the occurrence of behavior problems through the reports of family members, registered nurses or licensed practical nurses (Wagner et al., 1995a, 1995b). The measure was modified in three ways: (a) reports of resident behavior problems were obtained from direct care staff (i.e., CNAs) who have the most contact with residents; (b) the frequency of resident behavior problems was reported in addition to the occurrence; and (c) we attempted to assess burden that might be experienced by staff (see Study 4).

We examined the factor structure of the frequency of behavior problems in this adapted scale and compared it with the factor structure obtained from the community version of the RMBPC. We also examined the internal consistency and reliability of behavior problem frequency in this modified measure: (a) test–retest stability with the same CNA at two time points, (b) interrater reliability to examine the extent of agreement using the same CNA's responses to 2 interviewers, and (c) interrater reliability to examine the extent of agreement between 2 CNAs with similar knowledge of the index resident's behavior problems at two time points. Examination of discrepancies between these two types of interrater reliability could provide information regarding the extent of rater bias related to different raters' unique perceptions of the same resident (Hoyt, 2000; Hoyt & Kerns, 1999). Finally, we examined the construct and criterion-related validity of behavior problem frequency items on the Revised Memory and Behavior Problems Checklist—Nursing Home (RMBPC–NH), comparing the subscales obtained from confirmatory factor analysis with other measures of similar constructs. Because these experimental questions were addressed using different available nursing home samples, we present them as four separate studies.

General Method

Participants

Participants were enrolled in intervention studies designed to improve communication between nursing home residents and staff (conducted at sites in Alabama and Florida; Allen-Burge, Burgio, Bourgeois, Sims, & Nunnikhoven, 2001; Bourgeois, Dijkstra, Burgio, & Allen-Burge, 2001; Burgio et al., 2001) or to reduce resident agitation (conducted at sites in Alabama; Burgio, 1997). Residents from both studies shared the following entry criteria: (a) age of at least 55 years; (b) absence of major sensory impairment; (c) life expectancy longer than 6 months; and (d) residence within the facility for at least 1 month. Residents differed

in that those enrolled in the communication skills study at Alabama and Florida sites were required to have a Mini-Mental Status Examination (MMSE; Folstein, Folstein, & McHugh, 1975) total score greater than 0 or Short Portable Mental Status Questionnaire (SPMSQ; Pfeiffer, 1975) with fewer than 10 errors, and retention of minimal ability in verbal communication involving spontaneous speech based on criteria established by Bourgeois (1993). For residents enrolled in the study of agitation in Alabama nursing homes, resident entry criteria included CNA report of clinically significant behavior problems.

Measurement

RMBPC–NH—This 42-item instrument was used to assess CNA report of the frequency of resident problem behavior during the past week. Frequency items were scaled from 0 (*not at all/not applicable*) to 5 (*several times a day*). Items from the scale were designed by Orr-Rainey and Teri (cited in Wagner et al., 1995a, 1995b) and Wagner et al. (1995a, 1995b) to assess the following categories of behavior: cognitive, emotional, functional, and disruptive. In both sites of the communication skills study (Allen-Burge et al., 2001; Bourgeois et al., 2001; Burgio et al., 2001), each problem behavior for which CNAs reported some frequency elicited two questions designed to measure staff burden. Burden items were scaled from 0 (*not at all*) to 4 (*extremely*). Examples of behavior problems in each frequency category as well as the two burden phrasings from the communication skills study can be found in Table 1. A comprehensive list of behavior problem frequency items can be found in Wagner et al. (1995a, 1995b) or obtained from the authors on request.

MMSE—The MMSE (Folstein et al., 1975), a measure of global cognitive ability, rates orientation, immediate and delayed recall for words, attention and concentration, language, and praxis (total score range = 0–30). The test–retest and interevaluator reliabilities are .89 and .83, respectively. In a sample of nursing home residents, the MMSE was found to be 81% sensitive and 83% specific, using a cutoff of 24 for significant cognitive impairment (Kafonek et al., 1989).

Activities of Daily Living (ADLs)—ADLs were assessed across the intervention studies by using one of two measures. The communication skills study used a version of the Functional Independence Measure (FIM)—REACH version (Gitlin et al., 2002; Hamilton, Laughlin, Fiedler, & Granger, 1994; Kidd et al., 1995), developed for the National Institute of Health funded cooperative (multisite) agreement: Resources for Enhancing Alzheimer's Caregiver Health (REACH; Coon, Schulz, & Ory, 1999). Unlike the REACH project, the entire Motor subscale consisting of sphincter control, transfer, and locomotion was used in addition to self-care items (total score range = 13–91). Reliability data indicate an intraclass correlation coefficient of .96 for the motor domain, with unweighted kappas ranging from .53 to .66. Raw score means indicated that residents in the communication skills study had moderate functional impairments in Alabama ($M = 47.78$, $SD = 23.10$; range = 13–89) and Florida ($M = 55.99$, $SD = 31.05$; range = 13–91).

The agitation study used the Physical Self-Maintenance Scale (PSMS; Lawton & Brody, 1969). Six ADL domains (i.e., toileting, feeding, dressing, grooming, locomotion, and bathing) were represented by a 5-point Likert-type scale (total score range = 6–30). The

measure has been shown to have satisfactory interrater reliability based on the independent ratings of nurses (Pearson's $r = .87$) and trained research assistants (Pearson's $r = .91$; Lawton & Brody, 1969). Lawton and Brody also demonstrated adequate construct validity of the PSMS based on its association with other measures of functional competence. Raw score means indicated that these residents had moderate to severe functional impairments ($M = 20.77$, $SD = 5.31$; range = 8–30).

Study 1: Confirmatory Factor Analysis

Our primary purpose in Study 1 was to modify the nursing home version of the MBPC (Wagner et al., 1995a, 1995b) by assessing the factor structure of behavior problem frequency in the modified scale. Data for this study were obtained from residents and staff at both sites of the communication skills study (Allen-Burge et al., 2001; Bourgeois et al., 2001; Burgio et al., 2001) and from the agitation study (Burgio, 1997).

Method

Settings—This study was conducted in 16 nursing homes in Alabama and Florida with an average census of 107 residents on three units within each nursing home. On average, 23% of the beds were designated Medicare, 32% were Medicaid, and the remainder were private pay. Residents had an average length of stay of 3 years. Resident-to-CNA ratios were 9:1 during the day shift and 11:1 during the evening shift. Yearly rates of staff turnover were 33% among registered nurses, 38% among licensed practical nurses, and 37% among CNAs.

Participants

Residents: The average age of residents ($N = 309$; Alabama communication, $n = 96$; Florida communication, $n = 84$; agitation, $n = 129$) was 83.87 years ($SD = 18.13$; range = 59–105); 77% were women, 83% were White, and 17% were African American. Their average MMSE score was 10.74 ($SD = 7.10$; range = 0–29). In order to assess residents' average ability to perform ADLs, raw scores were transformed into z scores where high numbers indicated greater independence. The average ADL z score for the entire sample was 0.01 ($SD = 0.96$; range = -1.45 – $+3.71$).

As expected, due to the different entry criteria in the communication skills and agitation studies, the residents differed significantly in cognitive status, $F(2, 297) = 18.11$, $p = .01$. Residents in the agitation project had significantly lower scores on the MMSE ($M = 8.33$, $SD = 7.10$) than did residents in the Alabama ($M = 13.91$, $SD = 6.42$) or Florida ($M = 10.95$, $SD = 6.46$) communication skills study. However, residents did not differ across sites or projects in age, gender, race, or functional status.

Staff: The CNA identified as the primary day or evening shift caregiver and observed to have primary responsibility for the completion of daily care for each nursing home resident was operationally defined as the staff caregiver ($N = 165$; Alabama communication, $n = 68$; Florida communication, $n = 36$; agitation, $n = 61$). Their average age was 38.66 years ($SD = 10.28$; range = 19–71). Ninety percent were women, 87% were African American, 12% were White, and 1% was either Native American or mixed racial/ethnic heritage. Average

education was 13.59 years ($SD = 1.23$; range = 11–17), and average experience as a CNA was 10.24 years ($SD = 7.06$; range = 0–38). The CNAs were similar in all relevant demographic characteristics across sites and projects (all $ps > .05$).

Results and Discussion

Behavior problem frequency items were scaled as follows: 0 (*not at all/not applicable*), 1 (*not in the past week*), 2 (*1 to 2 times in the past week*), 3 (*3 to 6 times in the past week*), 4 (*1 to 2 times per day*), or 5 (*several times per day*). Within the subscales proposed a priori for the MBPC–NH (Wagner et al., 1995a, 1995b), 44% of the total of cognitive items were coded as not occurring at all or not occurring in the past week, 43% of the total of emotional items were coded as not occurring at all or not occurring in the past week, and 69% of the total of functional or disruptive items were coded as not occurring at all or not occurring in the past week.

Table 2 shows the fit statistics for three confirmatory factor analysis models using LISREL 8 (Jöreskog & Sörbom, 1996a). The fit statistics of one-factor, three-factor, and four-factor models were examined and compared. The three-factor model was designed to replicate the factor structure found for the community RMBPC (i.e., Memory-Related Problems, Depression, and Disruption; Teri et al., 1992). The four-factor model was designed to replicate the subscales proposed a priori for the previous version of the MBPC–NH measuring occurrence of behavior problems without attempting to measure staff burden (Wagner et al., 1995a, 1995b).

In multifactor models, behavior problem frequency items were constrained to have zero loadings on all factors except the one thought to be responsible for its observed variance. Correlations between factors were allowed to be freely estimated. Two different estimation methods were examined. In one set of analyses, maximum-likelihood estimation was used to model the standard correlation matrix. In a second set of analyses, unweighted least squares estimation was used to model the polychoric correlation matrix. This method was employed due to the ordinal and skewed nature of the item data. Although weighted least squares estimation used with the polychoric correlation matrix is sometimes recommended when items with ordinal, but not interval, properties are the focus of analysis (Jöreskog & Sörbom, 1996b), the 42-item RMBPC–NH matrix was too large for the models to converge using this procedure with a sample size of 309 residents.

Using the traditional maximum-likelihood estimation for the Pearson's correlation matrix, the three-factor model provided moderate fit to the data, $X^2(816, N = 309) = 2,035, p < .01$; goodness-of-fit index (GFI) = .75, adjusted GFI (AGFI) = .72, normed fit index (NFI) = .52, RMSEA = .07. For GFI, AGFI, and NFI, possible values range from 0 to 1, with higher values indicating better model fit. For the RMSEA, lower values indicate better fit. Factor intercorrelations ranged from .69 to .70. Nested comparisons of the chi-square statistics indicated statistically significant improvements in fit for the three-factor model over the one-factor model, $X^2(3, N = 309) = 235, p < .01$. Nested comparisons of the chi-square statistics failed to show improvement for the four-factor model over the three-factor model.

The results of the confirmatory factor analysis using the unweighted least squares estimation for the polychoric correlation matrix were improved. The three-factor model provided moderate fit to the data, $X^2(816, N = 309) = 2,830, p < .01$; GFI = .90, AGFI = .89, NFI = .87, RMSEA = .09. Factor intercorrelations ranged from .71 to .78. Nested comparisons of the chi-square statistics indicated statistically significant improvements in fit for the three-factor model over the one-factor model, $X^2(3, N = 309) = 380, p < .01$. Again, nested comparisons of the chi-square statistics failed to show improvement for the four-factor model over the three-factor model.

For clinical and conceptual reasons, we recommend use of the three-factor model, replicating the factor structure of the community measure (Teri et al., 1992) rather than the rationally constructed MBPC–NH with four subscales (Wagner et al., 1995a, 1995b). In the nursing home version of the RMBPC, problems in functional ability loaded onto the Disruption factor. Also, we have chosen to rename the Depression subscale as Emotional to reflect the diversity of emotional content of some of the items (e.g., “How often has your resident heard voices or sounds when there was no voice or sound?”). On the RMBPC–NH, the Memory-Related Problems subscale has 12 items, the Emotional subscale has 12 items, and the Disruption subscale has 18 items, including items measuring problems in functional ability.

Study 2: Reliability

The primary aim in this study was to examine the internal consistency and reliability of the behavior problem frequency items of the RMBPC–NH, including test–retest stability and interrater reliability (see Table 3). The reliability sample represents a subset of the total sample in Study 1. Specifically, the reliability sample consists of the participants in the communication skills study conducted in Alabama nursing homes (Burgio et al., 2001).

Method

Settings—This study was conducted in five nursing homes in the Birmingham metropolitan area with an average census of 120 residents on three units within each nursing home. On average, 11% of the beds were designated Medicare, 50% were Medicaid, and the remainder were private pay. Residents had an average length of stay of 3 years. Resident-to-CNA ratios were 9:1 during the day shift and 12:1 during the evening shift. Reported yearly rates of staff turnover were 8% among registered nurses, 18% among licensed practical nurses, and 49% among CNAs.

Participants

Residents: The average age of residents ($n = 94$) was 83.57 ($SD = 8.59$; range = 61–105); 74% were women, 83% were White, and 17% were African American. Their average MMSE score was 13.91 ($SD = 6.42$; range = 1–29), significantly higher than in the Florida site ($M = 10.95, SD = 6.46$) or in the agitation project ($M = 8.33, SD = 7.10$). Their average ability to perform ADLs using a z-score transformation was 0.05 ($SD = 1.00$; range = –1.45–1.84).

Staff: The CNA identified by the registered nurse and observed to have primary responsibility for the completion of daily care for each nursing home resident was operationally defined as the primary staff caregiver ($n = 68$). Their average age was 37.93 years ($SD = 10.26$; range = 19–60). Ninety-two percent were women, 87% were African American, and 13% were White. Their average education was 13.77 years ($SD = 1.12$; range = 11–15) and their average experience as a CNA was 9.58 years ($SD = 5.98$; range = 0–25). The CNAs did not differ from the larger sample in relevant demographic characteristics (all $ps > .05$).

Procedure—During a 4-week period designed to assess resident behavior in the nursing home, the RMBPC–NH was administered once per week (see Table 3). In Week 1, the primary day or evening shift CNA responsible for the resident's care was asked about the frequency of any problem behaviors displayed by the resident in the past week. In Week 2, the same interviewer asked the same CNA about behavior problems displayed by the same resident. By comparing responses in Week 2 with responses in Week 1, test–retest stability could be assessed. In Week 3, a different interviewer asked the same CNA about behavior problems displayed by the same resident. Correlating CNA responses to frequency items in Week 1 with Week 3 when a different research assistant interviewed the same CNA yielded an estimate of the interrater reliability of the assessors. In Week 4, the original interviewer asked a second CNA (termed *secondary CNA*) from the same shift with similar familiarity with the resident about the resident's behavior problems in the past week. By comparing responses in Week 4 with Week 1, interrater reliability of behavior ratings by 2 CNAs with similar familiarity with the resident could be assessed.

Although assessing the stability and reliability of survey data across a 4-week study period is not ideal, the number of assessments involving CNAs in the communication skills study precluded a shorter assessment interval. Given that the reliability of the survey was assessed across a 4-week period, it was necessary to demonstrate that the residents' behavior did not, in fact, change during that time interval. We compared four domains of resident behavior across weeks using the computer-assisted behavior observation system (CABOS) described by Burgio et al. (2001). Average kappa reliability across all categories was .79 (range = .68–.95). Results indicated no change in residents' verbal behaviors, activity patterns, disruptive behaviors, or time spent restrained or not restrained across weeks. Previous studies also have shown that test–retest scores from staff-report measures are stable up to 1 month between assessments (Koss et al., 1997; Patterson et al., 1997).

Results and Discussion

The internal consistency of the three subscales and for the total measure was high (Memory-Related, $\alpha = .87$; Emotional, $\alpha = .78$; Disruption, $\alpha = .87$; total, $\alpha = .95$). Correlating CNA responses to frequency items in Week 1 with Week 2 yielded an estimate of test–retest reliability. In general, test–retest reliabilities for the three subscales and for the total measure were high (Memory-Related, $r = .77$; Emotional, $r = .88$; Disruption, $r = .78$; total, $r = .86$). Correlating CNA responses to frequency items in Week 1 with Week 3 when a different research assistant interviewed the CNA yielded an estimate of interrater reliability between assessors that also was quite high for the three subscales and for the total measure (Memory-

Related, $r = .83$; Emotional, $r = .85$; Disruption, $r = .78$; total, $r = .88$). This suggests that (a) assessors can be trained to administer the measure in a similar manner; and (b) report of residents' behavior across a 3-week time interval is relatively stable.

The primary and secondary CNAs' familiarity with the resident was examined using a $2 \text{ (CNA)} \times 5 \text{ (nursing home site)}$ analysis of variance (ANOVA) where the outcome was the number of days the primary and secondary CNA worked on the resident's unit. Only the main effect of site was significant, $F(4, 95) = 4.29$, $p = .01$, indicating that CNAs worked a variable number of days across nursing homes, but that primary and secondary CNAs for a given resident worked a comparable number of days on the resident's unit. Comparing the responses from the primary CNA in Week 1 with a second CNA in Week 4 with similar familiarity with the resident revealed only moderate concordance regarding the frequency of resident behavior problems (Memory-Related, $r = .45$; Emotional, $r = .53$; Disruption, $r = .43$; total, $r = .46$). Considering the stability of behavior problems established with the CABOS and with test-retest and interrater reliabilities among assessors on the RMBPC-NH, it would appear that low concordance in CNAs' report of resident behaviors across the 4-week period examined did not reflect actual changes in the frequency of resident problem behaviors. These results suggested that the RMBPC-NH is an internally consistent and reliable indicator of the frequency of behavior problems among residents. However, CNA perception significantly impacted the report of behavior problem frequency.

Study 3: Construct Validity

We examined the construct (i.e., convergent and divergent validity) and criterion-related validity of the RMBPC-NH, comparing the total frequency of behavior problems with other measures of similar and dissimilar constructs in Study 2 and examining the frequency of behavior problems across differing subsamples from the communication skills study and agitation study described in Study 1.

Method

Measurement—Measures of cognitive (i.e., MMSE) and functional status (i.e., z-score transformation of ADL measures) have been previously described.

Geriatric Depression Scale (GDS): The GDS (Brink et al., 1982; Yesavage et al., 1982) is a 30-item depression screening inventory that uses dichotomous response options to indicate the presence or absence of depressive symptoms. Reliability data indicate that above an MMSE of 14, the GDS has a sensitivity of 84% and specificity of 91% (McGivney, Mulvihill, & Taylor, 1994). Due to concerns about the validity of self-report in this sample, this instrument was administered to CNAs well acquainted with the resident (Logsdon & Teri, 1995).

Occupational Stress Inventory (OSI): The OSI (Osipow & Spokane, 1983) is a 60-item questionnaire that was administered to CNAs in interview format to determine the amount of job stress they were experiencing. The standardized alpha coefficient for the total questionnaire is .93.

Results and Discussion

Convergent and Divergent Construct Validity—Regarding associations between the frequency of resident behavior problems and other staff and resident characteristics, a nested model analysis was conducted in Study 2 with residents nested within CNA respondents (Raudenbush, Bryk, Cheong, & Congdon, 2001). This was done in order to remove explained variance from the error term associated with 1 CNA respondent who reported the behavior problem frequency of more than 1 resident. There were no staff-level predictors for total frequency of resident behavior problems using the final estimation of fixed effects of the variables examined (i.e., work experience, total staff-reported job stress).

Regarding resident-level predictors, the *z*-score transformation of residents' ability to perform ADLs was significant, $t(290) = -2.60$, $r' = -.14$, $p = .01$, as was the residents' MMSE score, $t(290) = -6.60$, $r' = -.34$, $p = .01$. CNAs reported more frequent behavior problems to be associated with poorer functional status and lower cognitive functioning. More frequent behavior problems were also associated with CNA report of greater resident depression, $t(86) = 3.69$, $r' = .35$, $p = .01$. Residents' age, gender, and race were not found to be significant predictors of the frequency of resident behavior problems reported by CNAs.

Criterion-Related Validity—Planned pairwise comparisons in Study 1 revealed significantly more total behavior problems reported in the agitation study, $F(2, 306) = 26.84$, $p = .01$. This is not surprising considering that residents were selected based on the presence of clinically significant agitated behavior. Residents in the agitation study displayed an average total behavior problem frequency score of 61.44 ($SD = 30.12$) in comparison with the Alabama ($M = 38.90$, $SD = 28.16$) or Florida ($M = 37.67$, $SD = 21.36$) communication studies. Additionally, total behavior problem frequency was significantly associated with observed verbal disruption using the CABOS system in the agitation study ($r = .30$, $p < .01$). Thus, the RMBPC–NH adequately measured differential frequencies of resident problem behavior relative to residents selected for a study based on the presence of clinically significant staff-reported and observed, real-time behavioral agitation.

Study 4: Feasibility of Measuring Burden Experienced by CNAs

As stated previously, none of the existing measures of behavior problems in nursing homes incorporate an assessment of the burden experienced by direct care staff as a result of resident behavior problems. The wording of the burden item on the community measure (Teri et al., 1992) assesses the level of “bother” experienced by the primary caregiver. We did not use this wording because CNAs did not know residents in a prior relationship, and thus may experience stress in a manner different from informal caregivers (Chappell & Novak, 1994).

As shown in Table 1, the definitions of burden used in both sites of the communication skills study were designed to measure care-related and social-emotional burden experienced by formal caregivers providing intimate physical care as part of their job responsibilities in the workplace. In the communication study, we found that CNAs reported burden associated with the impact of behavior problems on providing care to the resident almost twice as

frequently as burden associated with the impact of behavior problems on the well-being of other residents and staff on the unit (see Table 1). Therefore, the assessment of staff's social-emotional burden was dropped from the agitation study and all analyses were done on care-related burden.

Method

Settings—The evaluation of care-related burden experienced by CNAs was conducted using only the resident and staff samples in the agitation study due to the greater frequency of problem behaviors in this sample. This study was conducted in six nursing homes in the Birmingham metropolitan area with an average census of 106 residents on an average of three units within each nursing home. Seven percent of the beds were designated Medicare, 31% were Medicaid, and the rest were private pay. Residents had an average length of stay of 3 years. Resident-to-CNA ratios were 9:1 during the day shift and 11:1 during the evening shift. Reported yearly rates of staff turnover were 41% among registered nurses, 48% among licensed practical nurses, and 51% among CNAs.

Participants

Residents: The average age of residents ($n = 129$) was 83.50 years ($SD = 8.62$, range = 59–103); 76% were women, 88% were White, and 12% were African American. Their average MMSE score was 8.33 ($SD = 7.10$; range = 0–26) and their average ability to perform ADLs, measured using a z-score transformation, was -0.01 ($SD = 0.91$; range = -1.16 – $+3.71$). As stated previously, these residents had lower MMSE scores in comparison with residents from the Alabama ($M = 13.91$, $SD = 6.42$) or Florida ($M = 10.95$, $SD = 6.46$) communication skills study.

Staff: The average age of CNAs ($n = 61$) was 40.77 years ($SD = 11.89$; range = 20–71); 90% were women, 84% were African American, and 16% were White. Their average education was 13.16 years ($SD = 1.40$; range = 11–17) and their average experience as a CNA was 10.76 years ($SD = 8.67$; range = 0–38). The CNAs did not differ from the larger sample in relevant demographic characteristics (all $ps > .05$).

Design and Procedures—Each of the 42 questions regarding the frequency of behavior problems among residents was associated with a question to assess possible care-related burden associated with the behavior. As shown in Table 1, the wording of the care-related burden item was “How much does this behavior affect your ability to give care to the resident?” Items were scaled from 0 (*not at all*) to 4 (*extremely*). If the CNA reported a behavior problem with a frequency of at least one to two times in the past week, the burden item was asked.

Results and Discussion

Descriptive data for the average number of overall behavior problems, total behavior problem frequency score, total burden score, average item frequency of behavior problems, and average burden when behavior problems were reported are displayed in Table 4. Of the 42 items on the RMBPC–NH, 100% of residents displayed at least 1 behavior problem per week. CNAs reported the occurrence of 16 ($SD = 7.48$; range = 1–32) behavior problems per

resident on average. On average, the observed frequency of these behavior problems ranged between three to six times per week and one to two times per day (see Table 4). The observed minimum frequency of behavior problems was reported to be one to two times per week, and the observed maximum was reported as several times per day.

The average total burden score was 9.42 ($SD = 18.69$; range = 0–90), but the median burden was 0 (see Table 4). Notably, 53% of residents had no care-related burden reported by CNAs associated with any behavior problems the resident displayed. For the 47% of residents ($n = 61$) who displayed behavior problems associated with care-related burden, staff reported themselves to be at least moderately burdened by the behavior ($M = 1.85$, $SD = 0.90$).

Regarding associations between reported care-related burden and CNA and resident characteristics, a nested model analysis was conducted with residents nested within CNA respondents (Raudenbush et al., 2001). Regarding staff-level predictors, no associations were found between report of total care-related burden and work experience as a CNA. Regarding resident-level predictors, no associations were found between report of total care-related burden and residents' age, cognitive status, or functional status. However, there were significant bivariate associations between total burden and more frequent resident behavior problems measured by the RMBPC–NH total score ($r = .49$, $p < .00$) and the total time the resident displayed disruptive vocalization measured by the CABOS ($r = .27$, $p < .01$).

General Discussion

This study evaluated the psychometric properties of a modified instrument for use in screening for behavior problems and associated burden among formal caregivers of nursing home residents with dementia across three distinct study samples. Results suggest that the RMBPC–NH is a reliable and valid tool for the assessment of the total frequency of behavior problems among nursing home residents and for the assessment of the frequency of behavior problems in three specific domains: Memory-Related, Emotional, and Disruption. We assert that the RMBPC–NH represents an improvement in the assessment of behavior problems exhibited by nursing home residents in comparison with other measures. The RMBPC–NH (a) measures the frequency of behavior problems in addition to the simple occurrence/nonoccurrence of these behaviors; (b) obtains reliable reports directly from CNAs, the staff members with the most direct contact with residents; (c) has a factor structure that mirrors that of the extensively used RMBPC community version; and (d) incorporates a measure of care-related burden experienced by formal caregivers (i.e., CNAs). Other commonly used staff-report measures of resident behavior problems do not attempt to assess staff burden. Furthermore, prior studies of the MBPC–NH did not conduct a confirmatory factor analysis and measured the occurrence, but not the frequency, of behavior problems (Wagner et al., 1995a, 1995b). Thus, we propose that the RMBPC–NH provides additional information with potential clinical utility in comparison with other measures.

We examined the factor structure of behavior problem frequency items across two geographically distinct samples from a communication skills training intervention (Allen-Burge et al., 2001; Bourgeois et al., 2001; Burgio et al., 2001) and from an intervention targeting agitation among demented nursing home residents (Burgio, 1997). Results

supported a three-factor model similar to that found with previous community samples (Teri et al., 1992), and a recently completed confirmatory factor analysis with the multisite, multiethnic REACH community sample (Roth et al., 2003). Our nursing home measure, however, incorporated greater diversity of items in the emotional domain; thus, we renamed the subscale Emotional, rather than Depression.

This series of studies also supported the criterion-related and the construct validity of the RMBPC–NH. CNA report of the frequency of resident behavior problems was significantly greater in the study in which residents were selected on the basis of exhibiting clinically significant agitated behavior. The construct validity of the RMBPC–NH, however, was not as strong as that observed for the community measure (Teri et al., 1992). Using hierarchical linear modeling, poor resident cognitive and functional status as well as more CNA-reported symptoms of resident depression were related to CNAs' report of more frequent behavior problems. This pattern of results potentially illustrates the CNAs' underlying perception of the resident, reflected in substantial correlations between staff-reported resident depression and resident behavior problems.

Our examination of reliability data was restricted to the Alabama communication skills sample. This sample of residents had significantly better cognitive status ($M = 13.91$, $SD = 6.42$) than either the Florida-based communication skills sample ($M = 10.95$, $SD = 6.46$) or the sample from the agitation study ($M = 8.33$, $SD = 7.10$). Results indicated high internal consistency for the total measure ($\alpha = .95$) and for the three subscales (α range = $.78$ – $.87$). Test–retest reliability ($r_s = .77$ – $.88$) was high, as was interrater reliability to examine the extent of agreement using the same CNA's responses to 2 interviewers ($r_s = .78$ – $.88$).

Perhaps the most significant finding in this study was the lack of agreement between 2 CNAs with similar familiarity with an index nursing home resident on the frequency of behavior problems exhibited by that resident. Specifically, interrater reliability assessed between 2 CNAs ranged from $.43$ to $.53$. Only one other study has assessed this type of reliability, showing low interrater reliability between “matrons” and “care assistants” rating the frequency and difficulty of behavior problems among residents in residential and nursing homes across England (Moniz-Cook, Woods, Gardiner, Silver, & Agar, 2001).

Given that (a) the measure itself was reliable as indicated by high internal consistency, test–retest stability, and interrater reliability across 2 assessors in a 3-week time interval; (b) primary and secondary CNAs had similar familiarity with the index resident and worked on the same shift; (c) residents' behaviors did not change across the 4-week period examined as measured by the CABOS; and (d) previous studies also have shown that test–retest scores from staff-report measures are stable up to 1 month between assessments (Koss et al., 1997; Patterson et al., 1997), we assert that the attenuated association between the behavior problem frequency reports of 2 CNAs can be attributed substantially to the CNAs' differential perceptions of resident behavior or different understanding of the items on the measure.

There are several explanations for this finding. Staff rating of the frequency of various target behaviors may strongly reflect the degree to which these behaviors capture their attention

and require their intervention or involvement (McCann, Gilley, Hebert, Beckett, & Evans, 1997). Such observer bias is not uncommon in the report of frequency of occurrence of different behaviors (Hoyt, 2000; Hoyt & Kerns, 1999). Notably, observer bias may partially account for the varied findings in the literature regarding predictors of resident agitation (Vance et al., 2003). Although CNAs are the staff members in nursing homes with the most frequent and intimate resident contact (Burgio et al., 1990; Chappell & Novak, 1994; Tellis-Nayak & Tellis-Nayak, 1989), they are also the most poorly educated and trained, and the lowest paid. Indeed, obtaining CNAs' responses to surveys regarding residents must be accomplished within the context of CNAs' provision of daily care and the many other distractions of a nursing home environment.

It is noteworthy that much of what is known nationally about quality-of-care practices in nursing homes and nursing home resident quality of life is based on staff ratings reflected in the federally mandated Minimum Data Set (MDS), and many authors have noted the limitations of MDS data, particularly in regard to behavioral agitation, pain report, and emotional functioning (Casten, Lawton, Parmelee, & Kleban, 1998; S. E. Fisher et al., 2002; Horgas & Margrett, 2001; Lawton et al., 1998; McCurren, 2002; Schnelle, Wood, Schnelle, & Simmons, 2001). Potential observer bias in CNA report of resident functioning also complicates the interpretation of information obtained from intervention studies in which staff-reported resident characteristics are a primary outcome. Interpretation of intervention outcomes may be compromised if outcome measures are administered to different CNAs, yielding potentially different results due to observer bias or different understanding of the items on the measure. In order to facilitate the accuracy and validity of data, intervention studies with staff-reported resident behavior as a primary outcome should consider: (a) averaging the report of two different CNAs with similar exposure to the resident in a manner similar to that reported by Moniz-Cook et al. (2001); (b) be careful to administer paper-and-pencil measures to the same CNA pre- and postintervention and to report the percentage of time this occurs (Burgio et al., 2001); or (c) train CNAs to rate resident behaviors more reliably.

To our knowledge, this study is the first attempt to evaluate burden experienced by CNAs associated with the frequency of behavior problems of residents under their care. Burden has long been associated with the care of individuals with dementia in the community (Teri et al., 1992; Zarit et al., 1980), and resident agitation has been associated with staff distress in nursing homes (Chappell & Novak, 1994; Everitt et al., 1991; J. E. Fisher et al., 1993). Because CNAs provide care to residents as part of their job, we did not use the traditional wording of "bother" used to assess burden among family caregivers in the community. Instead, we examined two different burden constructs: care-related burden and social-emotional burden (see Table 1) in the Alabama and Florida communication skills study (Allen-Burge et al., 2001; Bourgeois et al., 2001; Burgio et al., 2001). We found that social-emotional burden was rarely reported. Thus, only care-related burden was assessed in the agitation study (Burgio, 1997).

In comparison with a recently completed psychometric evaluation of the REACH version of the community RMBPC (Roth et al., 2003), CNAs reported a comparable number of behavior problems per resident (i.e., in terms of proportion of possible items: $M = 15.96$, SD

= 7.48, 38% of possible items, nursing home; $M = 10.22$, $SD = 4.16$, 43% of possible items, community). However, a large majority of community caregivers (93%) reported some burden in the REACH data, whereas only 47% of nursing home residents were reported to have burdensome problem behaviors by their formal, CNA caregivers. The total burden score was less in the nursing home sample than in the REACH community sample ($M = 9.42$, $SD = 18.69$, nursing home; $M = 16.73$, $SD = 13.64$, community). Notably, when burden was reported by CNAs, the average burden per item ($M = 1.85$, $SD = 0.90$) was larger than that reported by familial, community caregivers ($M = 1.49$, $SD = 0.91$; Roth et al., 2003). The data reflect moderate burden experienced by both familial and formal caregivers associated with behavior problems among older adults with dementia. We assert that having a measure of burden experienced by formal CNA caregivers of nursing home residents could be useful in designing interventions to decrease CNA turnover and to improve the quality of care of nursing home residents. Furthermore, a measure of CNA burden associated with resident problem behaviors could inform the design of staff-implemented resident interventions and staff motivational systems.

As with any empirical research, the limitations of this study must be acknowledged. First, the turnover rate in our nursing homes was about half the national norms (i.e., 97% for CNAs as cited in Harrington et al., 2000). Large staff turnover rates present significant difficulties in implementing nursing home research. Certainly, staff turnover presents a greater challenge for psychosocial interventions that might involve training CNAs to rate resident behaviors more reliably, and for studies conducted across greater periods of time than our maximum 4-week interval (in this regard, it is notable that MDS data are gathered once per quarter). Second, examination of construct validity was limited by discrepancy among the three projects in regard to ancillary survey data available for comparison. Third, examination of burden was potentially constrained by the wording of the burden questions. Although the CNAs in this study did respond to the care-related burden question, an alternative wording might be more effective in eliciting burden reports. We encourage researchers to investigate other facets of burden that may be more relevant to the CNAs' day-to-day activities.

In conclusion, the RMBPC–NH represents a reliable and valid assessment of the frequency of nursing home residents' behavior problems as reported by CNAs in three domains: Memory-Related, Emotional, and Disruption. The assessment of burden among CNAs should be further explored, although, in this study when CNAs reported burden, their report was consistent with the report of the burden experienced by family caregivers in the community. Test–retest stability (2-week interval) and interrater reliability (3-week interval) of the RMBPC–NH across assessors were adequate. However, interrater reliability across 2 CNAs with similar familiarity with an index nursing home resident was modest and suggests marked observer bias or differential interpretation of the items on the measure. This finding has clinical implications for the interpretation of MDS data and the design of research conducted in nursing homes, particularly for the evaluation of psychosocial interventions.

Acknowledgments

Funding from the National Institute on Aging (Grant R01AG13008) to Michelle S. Bourgeois and Louis D. Burgio and the National Institute of Nursing Research (Grant R01NR02988) to Louis D. Burgio supported the research reported here. Funding from the National Institute on Aging (Grant K01AG00943) to Rebecca S. Allen supported preparation of this article.

We thank the nurses, nursing assistants, and administrative staff of the many Birmingham and Tallahassee area nursing homes for their support and assistance. Without their cooperation these projects would not have been possible. Special thanks are extended to the staff of the Center for Mental Health and Aging for assistance in data collection and manuscript preparation.

References

- Algase D, Beck C, Kolanowski A, Whall A, Berent S, Richards K, et al. Need-driven dementia-compromised behavior: An alternative view of disruptive behavior. *American Journal of Alzheimer's Disease*. 1996; 11:10–19.
- Allen-Burge R, Burgio LD, Bourgeois MS, Sims R, Nunnikhoven J. Increasing communication among nursing home residents. *Journal of Clinical Geropsychology*. 2001; 7:213–230.
- American Health Care Association. Facts and trends 1999: The nursing facility sourcebook. Author; Washington, DC: 1999.
- Beck C, Vogelpohl TS. Problematic vocalizations in institutionalized individuals with dementia. *Journal of Gerontological Nursing*. 1999; 25:17–26.
- Bourgeois MS. Effects of memory aids on the dyadic conversations of individuals with dementia. *Journal of Applied Behavior Analysis*. 1993; 26:77–87. [PubMed: 8473260]
- Bourgeois MS, Dijkstra K, Burgio L, Allen-Burge R. Memory aids as an augmentative and alternative communication strategy for nursing home residents with dementia. *AAC Augmentative and Alternative Communication*. 2001; 17:196–210.
- Brink TL, Yesavage JA, Lum O, Hiersema P, Adey N, Rose TL. Screening tests for geriatric depression. *Clinical Gerontologist*. 1982; 1:37–44.
- Burgio, LD., Principle Investigator. Audiotape intervention for agitation (National Institutes of Health Grant R01NR02988). National Institute of Nursing Research; Washington, DC: 1997.
- Burgio LD, Allen-Burge R, Roth DL, Bourgeois MS, Dijkstra K, Gerstle J. Come talk with me: Improving communication with nursing home residents during care routines. *The Gerontologist*. 2001; 41:449–460. [PubMed: 11490043]
- Burgio LD, Engel BT, Hawkins A, McCormick K, Scheve A. A descriptive analysis of nursing staff behaviors in a teaching nursing home: Differences among CNAs, LPNs, and RNs. *Gerontologist*. 1990; 30(1):107–112.
- Casten R, Lawton MP, Parmelee PA, Kleban MH. Psychometric characteristics of the Minimum Data Set I: Confirmatory factor analysis. *Journal of the American Geriatrics Society*. 1998; 46:726–735. [PubMed: 9625189]
- Chappell NL, Novak M. Caring for institutionalized elders: Stress among nursing assistants. *Journal of Applied Gerontology*. 1994; 13(3):299–315.
- Cohen-Mansfield J. Reflections on the assessment of behavior in nursing home residents. *Alzheimer's Disease and Associated Disorders*. 1994; 8(1):S217–S222.
- Cohen-Mansfield J. Assessment of disruptive behavior/agitation in the elderly: Function, methods, and difficulties. *Journal of Geriatric Psychiatry and Neurology*. 1995; 8(1):52–60. [PubMed: 7710649]
- Cohen-Mansfield J, Billig N. Agitated behaviors in the elderly: I. A conceptual review. *Journal of the American Geriatrics Society*. 1986; 34:711–721. [PubMed: 3531296]
- Cohen-Mansfield J, Marx MS, Rosenthal AS. A description of agitation in a nursing home. *Journal of Gerontology*. 1989; 44(3):M77–M84. [PubMed: 2715584]
- Coon, DW., Schulz, R., Ory, MG. Innovative intervention approaches for Alzheimer's disease caregivers. In: Beigel, D., Blum, A., editors. *Innovations in practice and service delivery across the lifespan*. Oxford; New York: 1999. p. 295-325.

- Davis LL, Buckwalter K, Burgio LD. Measuring problem behaviors in dementia: Developing a methodological agenda. *Advances in Nursing Science*. 1997; 20(1):40–55. [PubMed: 9266016]
- Deutsch LH, Rovner BW. Agitation and other noncognitive abnormalities in Alzheimer's disease. *Psychiatric Clinics of North America*. 1991; 14(2):341–351. [PubMed: 1676506]
- Everitt DE, Fields DR, Soumerai SS, Avorn J. Resident behavior and staff distress in the nursing home. *Journal of the American Geriatrics Society*. 1991; 39:792–798. [PubMed: 2071810]
- Fisher JE, Fink CM, Loomis CC. Frequency and management difficulty of behavioral problems among dementia patients in long-term care facilities. *Clinical Gerontologist*. 1993; 13(1):3–12.
- Fisher SE, Burgio LD, Thorn BE, Allen-Burge R, Gerstle J, Roth DL, Allen SJ. Pain assessment and management in cognitively impaired nursing home residents: Association of certified nursing assistant pain report, Minimum Data Set pain report, and anal-gesic medication use. *Journal of the American Geriatrics Society*. 2002; 50(1):152–156. [PubMed: 12028260]
- Folstein MF, Folstein SE, McHugh PR. Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*. 1975; 12(3):189–198. [PubMed: 1202204]
- Gitlin LN, Roth DL, Burgio L, Loewenstein DA, Winter L, Nichols L, et al. Assessment of dependence in individuals with Alzheimer's disease and dependence-associated burden: Psychometric evaluation of a new measure for use with caregiver–care recipient dyads. 2002 Manuscript submitted for publication.
- Hamilton BB, Laughlin JA, Fiedler RC, Granger CV. Interrater reliability of the 7-level functional independence measure (FIM). *Scandinavian Journal of Rehabilitation Medicine*. 1994; 26(3):115–119. [PubMed: 7801060]
- Harrington C, Kovner C, Mezey M, Kayser-Jones J, Burger S, Mohler M. Experts recommend minimum nurse staffing standards for nursing facilities in the United States. *Gerontologist*. 2000; 40(1):5–16. [PubMed: 10750309]
- Horgas AL, Margrett JA. Measuring behavioral and mood disruptions in nursing home residents using the Minimum Data Set. *Outcomes Management and Nursing Practice*. 2001; 5(1):28–35.
- Hoyt WT. Rater bias in psychological research: When is it a problem and what can we do about it? *Psychological Methods*. 2000; 5(1):64–86. [PubMed: 10937323]
- Hoyt WT, Kerns MD. Magnitude and moderators of bias in observer ratings: A meta-analysis. *Psychological Methods*. 1999; 4(4):403–424.
- Jöreskog, KG., Sörbom, DG. LISREL 8: User's reference guide. Scientific Software International; Chicago: 1996a.
- Jöreskog, KG., Sörbom, DG. PRELIS 2: User's reference guide. Scientific Software International; Chicago: 1996b.
- Kafonek S, Ettinger WH, Roca R, Kittner S, Taylor N, German PS. Instruments for screening for depression and dementia in a long-term care facility. *Journal of the American Geriatrics Society*. 1989; 37(1):29–34. [PubMed: 2642498]
- Kidd D, Stewart G, Baldry J, Johnson J, Rossiter D, Petruckevitch A. The Functional Independence Measure: A comparative validity and reliability study. *Disability and Research*. 1995; 17(1):10–14.
- Koss E, Weiner M, Ernesto C, Cohen-Mansfield J, Ferris SH, Grundman M, et al. Assessing patterns of agitation in Alzheimer's disease patients with the Cohen-Mansfield Agitation Inventory. The Alzheimer's Disease Cooperative Study. *Alzheimer's Disease and Associated Disorders*. 1997; 11:S45–S50.
- Lawton, MP. Behavior-relevant ecological factors.. In: Schaie, KW., Schooler, C., editors. *Social structure and aging: Psychological processes*. Erlbaum; Hillsdale, NJ: 1989. p. 57-58.
- Lawton MP, Brody EM. Assessment of older people: Self-maintaining and instrumental activities of daily living. *Gerontologist*. 1969; 9:179–186.
- Lawton MP, Casten R, Parmelee PA, Van Haitsma K, Corn J, Kleban MH. Psychometric characteristics of the Minimum Data Set II: Validity. *Journal of the American Geriatrics Society*. 1998; 46:736–744. [PubMed: 9625190]
- Lawton MP, Kleban MH, Moss M, Rovine M, Glicksman A. Measuring caregiving appraisal. *Journal of Gerontology*. 1989; 44:P91–P71. [PubMed: 2715591]

- Logsdon RG, Teri L. Depression in Alzheimer's disease patients: Caregivers as surrogate reporters. *Journal of the American Geriatrics Society*. 1995; 43(2):150–155. [PubMed: 7836639]
- McCann JJ, Gilley DW, Hebert LE, Beckett LA, Evans DA. Concordance between observation and staff rating of behavior in nursing home residents with Alzheimer's disease. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*. 1997; 52(2):63–72.
- McCurren C. Assessment for depression among nursing home elders: Evaluation of the MDS mood assessment. *Geriatric Nursing*. 2002; 23(2):103–108. [PubMed: 11956523]
- McGivney SA, Mulvihill M, Taylor B. Validating the GDS depression screen in the nursing home. *Journal of the American Geriatrics Society*. 1994; 42(5):490–492. [PubMed: 8176142]
- Moniz-Cook E, Woods R, Gardiner E, Silver M, Agar S. The Challenging Behaviour Scale (CBS): Development of a scale for staff caring for older people in residential and nursing homes. *British Journal of Clinical Psychology*. 2001; 40:309–322. [PubMed: 11593958]
- Osipow, SH., Spokane, AR. A manual for measures of occupational stress, strain, and coping: (Form E-2). Marathon Consulting and Press; Columbus, Ohio: 1983.
- Patterson MB, Mack JL, Mackell JA, Thomas R, Tariot P, Weiner M, Whitehouse PJ. A longitudinal study of behavioral pathology across five levels of dementia severity in Alzheimer's disease: The CERAD Behavior Rating Scale for Dementia. The Alzheimer's Disease Cooperative Study. *Alzheimer's Disease and Associated Disorders*. 1997; 11:S40–S44.
- Pfeiffer E. A Short Portable Mental Status Questionnaire for the assessment of organic brain deficit in elderly patients. *Journal of the American Geriatrics Society*. 1975; 23(10):433–441. [PubMed: 1159263]
- Raudenbush, SW., Bryk, AS., Cheong, YF., Congdon, R. HLM 5: Hierarchical linear and nonlinear modeling. Scientific Software International; Chicago: 2001.
- Roth DL, Burgio LD, Gitlin LN, Gallagher-Thompson D, Coon DW, Belle SH, et al. Psychometric analysis of the Revised Memory and Behavior Problems Checklist: Factor structure of occurrence and reaction ratings. *Psychology and Aging*. 2003; 18:906–915. [PubMed: 14692875]
- Ryden MB, Bossenmaier M, McLachlan C. Aggressive behavior in cognitively impaired nursing home residents. *Research in Nursing and Health*. 1991; 14(2):87–95. [PubMed: 2047539]
- Schnelle JF, Wood S, Schnelle ER, Simmons SF. Measurement sensitivity and the Minimum Data Set depression quality indicator. *Gerontologist*. 2001; 41:401–405. [PubMed: 11405438]
- Tellis-Nayak V, Tellis-Nayak M. Quality of care and the burden of two cultures: When the world of the nurse's aide enters the world of the nursing home. *Gerontologist*. 1989; 29:307–313. [PubMed: 2759451]
- Teri L, Truax P, Logsdon R, Uomoto J, Zarit S, Vitaliano PP. Assessment of behavioral problems in dementia: The Revised Memory and Behavior Problems Checklist. *Psychology and Aging*. 1992; 7:622–631. [PubMed: 1466831]
- Vance DE, Burgio LD, Roth DL, Stevens AB, Fairchild K, Yurick A. Predictors of agitation in nursing home residents. *Journal of Gerontology: Psychological Sciences*. 2003; 58B:P129–P137.
- Vitaliano PP, Russo J, Young HM, Teri L, Maiuro RD. Predictors of burden in spouse caregivers of individuals with Alzheimer's disease. *Psychology of Aging*. 1991; 6:392–402.
- Wagner AW, Teri L, Orr-Rainey N. Behavior problems among dementia residents in special care units: Changes over time. *Journal of the American Geriatrics Society*. 1995a; 43:784–787. [PubMed: 7602032]
- Wagner AW, Teri L, Orr-Rainey N. Behavior problems of residents with dementia in special care units. *Alzheimer's Disease and Associated Disorders*. 1995b; 9(3):121–127.
- Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M. Development and validation of a geriatric depression screening scale: A preliminary report. *Journal of Psychiatric Research*. 1982; 17(1):37–49. [PubMed: 7183759]
- Zarit SH, Reever KE, Bach-Peterson J. Relatives of the impaired elderly: Correlates of feelings of burden. *Gerontologist*. 1980; 20:649–655. [PubMed: 7203086]
- Zarit, SH., Zarit, JM. Cognitive impairment.. In: Lewinsohn, PM., Teri, L., editors. *Clinical geropsychology*. Pergamon Press; New York: 1983. p. 38–81.

Table 1

Exemplar Behavior Problem Frequency Items by Subscale and Two Types of Burden

Category	Wording
Behavior problem frequency questions	
Memory-related	"How often has your resident been unable to follow verbal directions?"
Emotional	"How often has your resident appeared sad or depressed?"
Disruption	"How often has your resident resisted bathing or personal hygiene?"
	"How often has your resident engaged in combative behavior (hitting, scratching, or biting)?"
Burden questions	
Care-related	"How much does this behavior affect your ability to give care to the resident?"
Social-emotional	"How much does this behavior affect the well-being of other residents and staff on the unit?"

Table 2

Fit Statistics for the Confirmatory Factor Analysis Models (N = 309)

Method and model	χ^2	df	GFI	AGFI	NFI	RMR	RMSEA
ML/R							
1-factor	2,270.1	819	.71	.68	.47	.08	.08
3-factor	2,035.3	816	.75	.72	.52	.08	.07
4-factor	2,032.0	813	.75	.72	.52	.08	.07
ULS/P							
1-factor	3,210.5	819	.89	.88	.86	.11	.10
3-factor	2,830.1	816	.90	.89	.87	.10	.09
4-factor	2,822.8	813	.90	.89	.87	.10	.09

Note. ML/R = maximum likelihood estimation for the Pearson's correlation matrix; ULS/P = unweighted least squares estimation for the polychoric correlation matrix; GFI = goodness-of-fit index; AGFI = adjusted goodness-of-fit index; NFI = normed fit index; RMR = root-mean-square residual; RMSEA = root-mean-square error of approximation.

Table 3

Graphic Representation of the Reliability Study

Week	Resident (R)	CNA (A)	Interviewer (I)
1	R ₁	A ₁	I ₁
2	R ₁	A ₁	I ₁
3 ^a	R ₁	A ₁	I ₂
4 ^b	R ₁	A ₂	I ₁

Note. CNA = certified nursing assistant.

^a In Week 3, a different interviewer administered the Revised Memory and Behavior Problem Checklist—Nursing Home (RMBPC–NH) to the same CNA regarding behavior problems exhibited by the index resident in the past week.

^b In Week 4, the original interviewer administered the RMBPC–NH to a different CNA from the same shift with equivalent familiarity with the index resident.

Table 4

Total Behavior Problem Frequencies and Associated Burden Data Averaged Across Items by Resident

Category	<i>n</i>	<i>M</i>	<i>Mdn</i>	<i>SD</i>	Obs. min	Obs. max	Possible range
Number of behavior problems in the past week	129	15.96	16.00	7.48	1	32	0–42
Total Frequency score	129	58.34	55.00	30.40	4	155	0–210
Total Burden score	129	9.42	0	18.69	0	90	0–168
Average frequency rating for behavior problems experienced in the past week	129	3.62	3.59	0.74	2	5	2–5
Average burden rating for behavior problems experienced in the past week	61	1.85	1.80	0.90	1	4	1–4

Note. Frequency of behavior problems were rated as follows: 0 (*never occurred*), 1 (*did not occur in the past week*), 2 (*1 to 2 times per week*), 3 (*3 to 6 times per week*), 4 (*1 to 2 times per day*), 5 (*several times per day*). Minimum (Obs. min.) and maximum (Obs. max.) observed values are reported for comparison with the possible range of values in each category. Only behavior problems occurring in the past week (i.e., frequency of 2–5) were included in calculations of the number of behavior problems. Care-related burden was rated as follows: 0 (*not at all*), 1 (*a little*), 2 (*moderately*), 3 (*very much*), 4 (*extremely*).