Measurement Sensitivity and the Minimum Data Set Depression Quality Indicator

John F. Schnelle, PhD,1,2 Stacey Wood, PhD,3 Elizabeth R. Schnelle, BS,3 and Sandra F. Simmons, PhD1

Purpose: The purpose of this study was to determine the accuracy of the prevalence rating of depression in nursing homes as flagged on the Minimum Data Set (MDS) quality indicator report. Design and Methods: Research staff measured depression symptoms and compared the results with the prevalence of disturbed mood symptoms documented by nursing home (NH) staff on the MDS in two samples of residents living in different NHs. The homes had been flagged on the nationally mandated MDS quality indicator report as having unusually low (Site 1) or high (Site 2) prevalence rates of depression. Results: The percentages of residents determined by research staff interview assessments to have probable depression in the two resident samples were not significantly different (49% vs. 55%, respectively) between homes. The staff in the home flagged on the MDS quality indicator report as having a high depression prevalence rate identified significantly more residents who also had scores indicative of probable depression on the resident interviews for follow-up mood assessments than did the home with a low quality indicator prevalence rate (78% vs. 25%, respectively). Implications: The prevalence of the depression quality indicator may be more reflective of measurement processes than of depression outcomes. Factors that may affect the difference in detection rates are discussed. Key Words: Quality indicators, Depression, Minimum Data Set

The development of quality indicators (QIs) based on Minimum Data Set (MDS) reports is a relatively recent and important innovation that should improve the effectiveness of quality assurance activities for all nursing homes (NHs; Arling, Karon, Sainfort, Zimmerman, & Ross, 1997). Currently, quarterly reports that compare individual NHs to all others within their respective state on 25 QIs are generated for all NHs in the United States. For example, the prevalence of the QI related to incontinence is adjusted into high and low risk groups on the basis of resident characteristics associated with incontinence (e.g., immobility, cognitive impairment). Individual NHs and survey staff are provided with a quarterly written report that provides a comparison of incontinence prevalence rates for all other facilities within the state. A facility that reports a 50% incontinence prevalence rate for low-risk residents, for example, might be placed in the 75th percentile in relation to other within-state facilities. This facility may, thus, be “flagged” on this QI with the consequence that state survey teams might pay particular attention to incontinence issues during their annual visits. Other QIs, such as the prevalence of depression, are not risk adjusted when comparisons are made against other facilities.

Preliminary work has described the stability and reliability of the MDS-based QIs despite the ac-
knowledge that more work needs to be done to understand what is being measured by the indicators (Karon, Sainfort, & Zimmerman, 1999; Zimmerman et al., 1995). More specifically, if the QIs are to be useful for both internal and external quality assurance purposes, then it is necessary to determine what factors differentiate facilities that score high and low on different indicators. There is a chance that even highly different percentile rankings on the QIs might not always reflect differences in care processes but rather differences in assessment processes. This article provides preliminary data relevant to this issue for the depression QI.

Methods

Participants and Setting

Participants were residents in two NHs in Southern California who were participating in a National Institutes of Health-sponsored clinical trial to evaluate the effects of an incontinence and exercise intervention. Site 1 was a 150-bed profit facility and Site 2 was a 200-bed nonprofit facility. There were two other important differences between the NHs that are directly relevant to the focus of this article. Site 2 was a university-affiliated facility and contracted for mental health services with the university. The mental health team frequently interacted with NH staff, maintained an “on-site” office, and attended many resident care meetings. The proprietary facility did not have an in-house mental health staff and, thus, referred all residents with psychiatric problems to off-site mental health staff. The second difference between the homes was that the proprietary facility (Site 1) had been flagged on a nationally mandated quality indicator report as having an unusually low prevalence rate of depression symptoms (1%), which placed them in the 1st percentile compared with all other California NH facilities. Site 2, the university-affiliated facility, was flagged as having an unusually high prevalence rate of depression symptoms on the same report (12%), which placed them in the 70th percentile compared with all other California NH facilities. The QI report covered the same time period that research staff directly assessed depression on residents. Site 1 was directly relevant to the focus of this study. Site 2 was included in the study to ensure that research staff directly assessed depression on residents following the beginning of the intervention trial. We obtained informed written consent, or the consent of a respective family member and the assent of the resident, from 148 (72%) of the eligible residents. We lost 18 consented residents prior to the beginning of the intervention trial for various reasons (e.g., death, transfer out of facility, consent withdrawal). We lost an additional 21 residents following the beginning of the intervention trial, primarily because of hospitalization or death.

Thus, 109 residents remained for participation in this study. We obtained all demographic and MDS information for the 109 participants. However, we only obtained depression interview information for 91 of these residents. Of the remaining 18 residents, 8 refused to cooperate with the interview, 6 were unable to participate in the interview (i.e., no response or nonsense responses to the interview questions), and 4 were not available for the interview (e.g., out of the facility, with family, in an activity).

Measures

We used two assessment instruments in this study. The Geriatric Depression Scale Short form (GDS-short) and the most recent MDS assessment, which is completed by the NH staff (Yesavage, Brink, Rose, & Lum, 1983). The 15-item short form of the GDS has been specifically recommended by one practice guideline as a screen for probable depression in NH residents (American Medical Directors Association, 1996). In this study, we administered the GDS-short to each participant in a one-on-one interview format because of the presence of cognitive and visual limitations among the NH residents in the samples. Despite the indication that the sensitivity and specificity of the GDS-short in detecting depression is reduced among individuals with dementia, researchers have documented that a total score greater than 5 on the GDS-short is related to probable depression (McGivney, Mulvihill, & Taylor, 1994).

In consideration of the possible influence of dementia on the reliability of GDS-short scores and to evaluate the stability of resident interview responses, we attempted to conduct a second GDS-short interview for a subsample of 11 residents in Site 1 (n = 33) and all residents in Site 2 (n = 58) who completed the first GDS-short interview. We attempted the second GDS-short interview on the day following the initial GDS-short assessment (i.e., two consecutive assessment days) at the same time of day (i.e., between 3 and 4 p.m.). Of the repeat interview attempts with 73 participants, nine residents in Site 1 and 39 residents in Site 2 (n = 48) provided complete interview information for the second GDS-short assessment. The remaining 25 residents either refused or were unavailable for a second interview (e.g., with family, in an activity, out of the facility).

We retrieved the MDS assessment (i.e., quarterly or annual) closest to the date of the GDS-short interview for each participant to assess NH staff documentation of the MDS items that trigger a follow-up assessment for a potential mood disturbance. These 16 MDS mood items are found in “Section E1: Indicators of Depression, Anxiety, and Sad Mood,” Items A through P (Health Care Financing Adminis-
tration, 1999). Finally, for the two samples of residents for whom GDS-short interview data were available, we generated the prevalence of the same depression QI that was generated by the state for all residents in the facility. MDS items that are used to generate the depression QI include the following: (a) Section E, Item 2 (sad mood); and (b) at least two symptoms of functional depression as measured by Items E1a (negative statements); E1g (suicidal or recurrent thoughts of death); E1n (repetitive physical movements); E1o (withdrawal from activities); E1p (reduced social activity); E1j (wakes with unpleasant mood), N1d (not awake most of the day), or N1a,b,c and B1 (awake one period of the day or less and not comatose); E4a (resists care); and K3a (weight loss).

Data Analyses

We compared all demographic and MDS information for those residents who provided complete interview data (n = 91) versus those who did not (n = 18). We conducted these same comparisons between those who participated in this study (n = 109) versus those who dropped out of the study (n = 39) and between the samples at each of the two NH sites. Finally, we also made a comparison between those who completed a GDS-short interview (n = 91) versus those who did not, for any reason (n = 39 + 18 = 57). We conducted independent sample t tests for all continuous variables (i.e., age, length of stay, Mini Mental State Exam [MMSE], total number of MDS symptoms of disturbed mood) and chi-square analyses for categorical variables (i.e., gender, ethnicity, presence or absence of a dementia or depression diagnosis, presence or absence of a “trigger” for disturbed mood). We conducted all statistical comparisons between homes with nonparametric analyses (chi-square). We analyzed the agreement between the two GDS-short assessments with Pearson product-moment correlation coefficients for GDS-short total scores and Kappa statistic for classification into two groups, those with GDS-short scores >5 (i.e., indicative of probable depression) and those with GDS-short scores ≤5 according to each of the two assessments.

Results

The mean age of the 91 participants was 89.1 (SD = 6.9). The participants were predominantly White (96%) and women (88%) with a mean length of stay of 26.8 (SD = 33.1) months. The mean MMSE score for the participants was 13.4 (SD = 7.7) with almost a full range of scores represented by the group (i.e., 0 to 29). Thirty-seven (41%) of the participants had a physician-recorded chart diagnosis of dementia. There were no significant differences on any of these descriptive characteristics between participants who completed a GDS-short interview (n = 91) and those who did not (n = 18), although four of the five residents who scored zero on the MMSE failed to complete the GDS-short interview. There were also no significant differences on any of these characteristics between all residents who either dropped out of the study or could not complete an interview (n = 39 + 18 = 57) and all residents who did complete a GDS-short interview (n = 91). Participants in Site 2 were significantly older than participants in Site 1 (M = 90.3, SD = 6.7 versus M = 86.1, SD = 6.3, respectively; t(127) = −3.67, p < .001). There were no other characteristics that were significantly different between the two sites.

Table 1 shows the comparisons between the GDS-short and MDS measures of mood for the two sites. The denominators in the cells of the table reflect the percentage provided in Table 1. The denominators vary between rows on the basis of the percentage statistic being calculated. For example, as illustrated in row 1 of Table 1, 49% of the residents in Site 1 and 55% in Site 2 who were able and willing to complete the GDS interview (n = 33 and n = 58, respectively, as reflected in the denominators) scored higher than 5 on the GDS-short, which is indicative of probable depression. There was no significant difference between sites on either the average GDS-short total scores (M = 6.4, SD = 3.7 vs. M = 5.3, SD = 3.9, respectively; t(89) = 1.3, p = .19) or the proportion of residents who scored higher than 5 (χ²(N = 91) = .38, p = .54). There was also no significant difference between sites in the proportion of residents who had probable depression according to the GDS-short (i.e., total score >5) when only those residents who did not have a chart diagnosis of dementia were considered (41% vs. 58%, χ²(N = 53) = 1.36, p = .243). In addition, the proportions of residents within the retest subsample who scored above or below the cut-off value of 5 on the GDS-short interview was also highly stable with 39 (81%) residents

<table>
<thead>
<tr>
<th>Measure</th>
<th>Site 1</th>
<th>Site 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>%</td>
<td>%</td>
<td>Denominator</td>
</tr>
<tr>
<td>GDS-short &gt;5</td>
<td>49</td>
<td>16</td>
</tr>
<tr>
<td>MDS Mood Symptom</td>
<td>34</td>
<td>13</td>
</tr>
<tr>
<td>MDS Mood Symptom and GDS &gt;5</td>
<td>25</td>
<td>4</td>
</tr>
<tr>
<td>Residents Depressed on MDS Quality Indicator</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

Note: GDS-short = Geriatric Depression Scale-short form; MDS = Minimum Data Set.

*Difference between sites p < .01.
receiving the same classification according to both assessments (26 scored >5 on both assessments; 13 scored ≤5 on both assessments). This comparison yielded a Kappa value of .60 (p < .001) which reflects good agreement between the two assessments. There was no significant difference between sites on test–retest agreement as determined by Kappa. The level of agreement continued to be significant despite a smaller sample when only those residents without a diagnosis of dementia were considered (Kappa = .48, p < .01). The correlation between the GDS-short total scores for the two assessments among the retest sample was .77 (p < .001).

Table 1, row 2 shows the number and percentage of residents in each NH who were rated by staff as having one or more (of a possible total of 16) symptoms of disturbed mood. There was a significant difference between NHs on this MDS mood measure; NH staff scored 34% of 38 participants in Site 1 and 72% of 71 participants in Site 2 as positive for a MDS mood symptom. Table 1, row 3, shows that NH staff in Site 2 also identified a higher proportion of residents with mood symptoms whom the research staff independently identified as expressing depression symptoms during the GDS-short interview as compared with Site 1 (78% vs. 25%, respectively; \( \chi^2 (N = 48) = 12.59, p < .001 \)).

Finally, the percentage within each of the two resident samples who would be identified as depressed on the basis of the depression QI calculation was significantly different and, in fact, reflected the differences in the depression QIs that were generated for the entire facility. Row 4 of Table 1 shows that the QI estimates of depression prevalence were 3% and 19% in the two samples, as compared with 1% and 12%, respectively, for the entire population in each facility.

In consideration of these findings, we conducted further analyses to clarify why NH staff in Site 1 detected less mood disturbance than NH staff in Site 2 as documented on the MDS. We first analyzed the number of residents in the two samples who were admitted to the NH with a diagnosis of depression on the premise that NH staff within each of the two homes might be differentially sensitized to disturbed mood symptoms based on this admission variable. We also conducted interviews with the MDS nurse at each site to determine the procedures used to complete the MDS items related to mood disturbance. We did not find differences between the NHs on either of these variables. The percentage of residents who were admitted to the NH with a diagnosis of depression in the two samples was not significantly different between sites (i.e., 28% in Site 1 and 31% in Site 2). Furthermore, both NHs employed an MDS nurse who did not have direct care responsibilities to complete the MDS. In both cases, these MDS nurses reported that they relied on direct care staff for information.

Because, according to GDS-short interview data, NH staff were not identifying symptoms among a substantial number of residents (12 at Site 1 and 7 at Site 2), who scored >5 on the GDS-short, we hypothesized that NH staff may be identifying that subset of residents who were expressing more severe or more noticeable symptoms of depression. Thus, we compared GDS-short total scores for those participants correctly identified by NH staff as having symptoms of disturbed mood versus those who were not identified by NH staff (i.e., GDS-short total score >5 but no MDS-documented symptoms of disturbed mood). However, we found the GDS-short total scores to be comparable between these two groups (\( M = 8.6, SD = 2.0 \) vs. \( M = 9.0, SD = 2.5 \), respectively).

To further assess which, if any, symptoms NH staff were more likely to document (i.e., notice) according to the MDS, we ran a frequency distribution of all sixteen mood items for the group of 109 participants for whom we had complete MDS information. This analysis showed the following symptoms to be documented most frequently among residents: (a) Item b, repetitive questions (\( n = 12 \)); (b) Item c, repetitive verbalizations (\( n = 12 \)); (c) Item d, persistent anger (\( n = 13 \)); (d) Item h, repetitive health complaints (\( n = 11 \)); (e) Item i, repetitive anxious non-health-related concerns (\( n = 11 \)); (f) Item l, sad and/or pained facial expression (\( n = 40 \)); (g) Item m, crying, tearfulness (\( n = 16 \)); (h) Item n, repetitive physical movements (\( n = 16 \)).

Those symptoms less frequently documented among residents by NH staff included the following: (a) Item a, negative statements (\( n = 3 \)); (b) Item e, self depreciation (\( n = 2 \)); (c) Item f, expression of unrealistic fears (\( n = 6 \)); (d) Item g, recurrent statements that something terrible is about to happen (\( n = 1 \)); (e) Item j, unpleasant mood in the morning (\( n = 3 \)); (f) Item k, insomnia and/or changes in sleep patterns (\( n = 6 \)); (g) Item o, withdrawal from activities of interest (\( n = 4 \)); and, (h) Item p, reduced social interaction (\( n = 7 \)).

**Discussion**

This study provides preliminary evidence that the prevalence of depression as measured by NH staff-generated estimates (i.e., MDS data) may be more influenced by the ability of indigenous nursing staff to detect symptoms than by the actual prevalence rate. It is unclear how the prevalence of the depression QI derived from the MDS will be interpreted or evaluated by survey staff or providers, who currently represent the primary users of this information, but it is clear that Site 2 should not be regarded as having a more serious problem with depression outcomes than Site 1. In fact, the opposite conclusion seems more defensible. There is a greater likelihood that Site 1 should be targeted for an intervention project to improve their ability to detect mood symptoms that may be indicative of depression. The fact that Site 2 had mental health staff who were frequently on site suggests that NH staff at Site 2 who completed the MDS were more aware of mood symptoms than the staff in Site 1 who did not have
an on-site mental health service. The fact that depression symptoms are generally underdetected in long-term care because providers regard the symptoms as a normal part of aging has been described in other reports (Katz, Streim, & Parmelee, 1994). These other reports, when considered in the context of the data reported in this article, suggest that training programs should be developed to improve depression symptom recognition among providers who do not have the unique advantage of having daily access to mental health professionals.

The data reported in this article are limited by the fact that they were collected as part of an ongoing clinical trial that was not directly designed to evaluate factors that might differentiate NHs on either depression outcomes or on the processes that NH staff use to detect symptoms of depression among residents. In addition to sample size and data generalizability limitations created by the study design, other issues could also be raised about our methodology to assess depression.

We have previously noted that there is some concern that the GDS-short has reduced sensitivity and specificity in detecting depression among individuals with dementia. No other standardized instrument, however, has been demonstrated to be a better depression screen, and there is expert consensus that the GDS-short should be used in the NH setting. Furthermore, the subsample of residents who completed two interviews in this study represented a range of cognitive impairment (i.e., those with and without a dementia diagnosis) and showed high reliability of both GDS total scores as well as classification as “probable depression” according to the GDS-short cut-off value of greater than 5. Finally, rates of probable depression on the basis of GDS-short interview results remained comparable between sites after controlling for dementia diagnoses.

It is important to recognize that both the GDS-short and the MDS should be considered only as screens for depressive symptoms that warrant further follow-up evaluation (as outlined in the MDS-Resident Assessment Protocols) and not as accurate measurements of a clinical diagnosis of major depressive disorder. A good case can, thus, be made from our data that the two homes differ significantly in their ability to screen for depressive symptoms. This could indicate a potential problem at least among that subsample of residents who participated in the clinical trial because it is likely that lack of identification translates into lack of appropriate treatment for these residents. These results suggest that the MDS QI report used to compare homes on multiple indicators might often be more influenced by differential assessment sensitivity as opposed to differential outcomes in at least one important care area.

References