Use of Proxy Respondents and Accuracy of Minimum Data Set Assessments of Activities of Daily Living

Terry Y. Lum,¹ Wen-Chieh Lin,² and Robert L. Kane³

¹University of Minnesota School of Social Work, St. Paul.
²Department of Family and Community Medicine, University of Missouri–Columbia.
³University of Minnesota School of Public Health, Minneapolis.

Background. Although the Minimum Data Set (MDS) presents a wide range of opportunities for policy makers and practitioners interested in outcomes of nursing home care for frail elderly persons, researchers have debated the validity and reliability of measurements in the MDS from the outset. To investigate this issue, the authors studied the accuracy of functional assessments by comparing the MDS and interview data collected in two evaluation studies.

Methods. Activities of daily living (ADL) assessment data from 3385 nursing home residents were collected from interviews with nursing home residents (n = 1200), family members (n = 1070), and nursing home staff (n = 1175). The MDS data for these nursing home residents were obtained and matched with the interview data. The agreement in ADL assessments between interview data and the MDS was assessed using Kappa statistics and multinomial logit regression for each of the three data sources.

Results. The agreement on ADL assessments between MDS and interview data was low to moderate (Kappa = 0.25 to 0.52), regardless of the sources of data. Interview data from staff and family proxies agreed to a greater degree with the MDS than did data collected from nursing home residents. The MDS reported fewer ADL difficulties than did staff proxies and more ADL difficulties than did nursing home residents. These findings held even after adjustment for other confounding factors using multinomial logit regression.

Conclusions. The substantial discrepancy between MDS and interview data can be attributed to both bias and error. The ADL assessments based on residents’ and family or staff reports differ, but the size of these differences depends on the proxy type and the method of data collection.

The Omnibus Budget Reconciliation Act of 1987 mandated the federal government to develop, implement, and evaluate a comprehensive system to assess all nursing home residents. The subsequent introduction of the Minimum Data Set (MDS) in 1991 was hailed as a major advance in nursing home quality improvement. All nursing homes certified by Medicare or Medicaid are required by law to complete a standardized MDS assessment of each resident’s cognitive, functional, health, and psychosocial status. The assessment is typically done by a registered nurse during a 7-day period. Assessments are required at admission, every 3 months after admission, and with any significant change in the resident’s status. The MDS provides a consistent set of clinical information on all nursing home residents that can be used to assess the care outcomes in nursing homes and compare the care outcomes across facilities and over time.

Although the MDS presents a wide range of potential opportunities for policy makers and practitioners interested in outcomes of nursing home care for frail elderly persons, researchers have debated the quality of measurements in the MDS since it was implemented (1–5). Some studies have evaluated the reliability and validity of MDS measurements in cognitive status (1,2,6,7), physical functioning (1,2,8,9), and pain (10,11), but many of them were conducted by the researchers who developed the MDS. Although the developers of the MDS suggested that accurate and reliable information could be obtained by standardizing nursing home...
residents, we need to recognize the possible bias introduced into the MDS data by the assessment methods.

Furthermore, the fact that the MDS assessment depends on information from different sources may also introduce random errors into the data. The accuracy of the MDS depends not only on whether the MDS coordinators would and could accurately judge and report the functioning of the residents but also on the communication among nursing home staff, thereby increasing the possibility that both bias and errors may be introduced into the assessment process.

The use of proxies in assessment has long been an integral part of studies of older adults. Proxy respondents provide supplemental information when an older adult has limited ability to participate in the data collection (19,20). There is a growing literature on the differences in health and disability assessments between respondents’ self-reports and those of proxies (19). In general, proxy respondents for older adults are likely to report higher rates of disability than are older respondents themselves. The level of discrepancy in reporting depends on how observable the disabilities are, how much interaction is needed to get help (21), and the level of cognitive impairment that the respondents have (19,22). The discrepancy between proxy assessment and respondent self-assessment is smaller for disabilities that are more observable and require more interaction to get help (21). In general, proxies are more likely to over-report functional limitations for older persons with dementia (19).

So far, most studies that describe the effect of proxy respondents on the quality of data collected have focused on community samples. Less research has been conducted on the quality of data relevant to physical functioning or self-care ability for a nursing home sample (20). We took advantage of an opportunity to compare ADL assessments in the MDS with ADL assessments gathered as part of two large evaluation studies in 6 states. The latter data were obtained from in-person interviews with nursing home residents, family members, and nursing home staff. This allowed us to identify the possible bias in ADL assessments introduced by the use of proxies in the MDS. We were especially interested in separating the effects of bias (based on the types of respondents) from errors.

METHODS

The goal of this study was to understand the sources of variation in ADL assessments in the MDS by comparing ADL data collected in the MDS with ADL data collected by trained interviewers through in-person interviews with nursing home residents, family members, and staff. The major difference in the MDS data and the interview data is the source of information. Data on ADLs in the MDS typically come from assessments by the MDS coordinators and other nursing home staff, whereas ADL data in the interview come from three possible sources: residents’ self-reports, staff proxy assessments, or family proxy assessments. By comparing the ADL data in the MDS with interview data from each of the three data sources, we sought to document the direction and magnitude of agreement and disagreement. Therefore, we hoped to shed new light on the quality of ADL data in the MDS. Based on the literature we reviewed, we expected that agreement would be highest between the MDS and interview data obtained from staff proxy and lowest between the MDS and interview data obtained from residents. Furthermore, we also expected that the MDS was more likely to report greater disability than interview data in the MDS/resident self-assessments comparison than in the other two (MDS/staff proxy assessments and MDS/family proxy assessments) comparisons.

Interview data for this study were originally collected for the evaluation of the EverCare program and the Minnesota Senior Health Care Option (MSHO) program. Information about the two evaluations is reported elsewhere in greater detail (23,24). The EverCare evaluation included an interview of 1269 nursing home residents in five states (Massachusetts, Florida, Georgia, Colorado, and Maryland): 441 were enrolled in the EverCare program and 828 were in two comparison groups. The study sample was chosen by the following procedure: All EverCare enrollees who volunteered for the evaluation were included in the enrollee sample. A matched sample of non–EverCare enrollees from the same nursing homes was selected based on a match using several characteristics (age, sex, race, previous nursing home residence before the start of EverCare, and their Medicare managed care enrollment status before EverCare enrollment).

A matched sample of non–EverCare enrollees in other nursing homes in the same geographic area was selected based on the best match in the same characteristics described in the previous paragraph. Trained interviewers interviewed the respondents. These interviewers had achieved a high level of inter-rater reliability before the actual interviews were conducted. Whenever possible, the residents were interviewed. When the residents were too frail to respond, the interviewers were instructed to interview a significant family member who had regular contact with the resident (family proxy). This significant family member was identified and confirmed with the residents. If no significant family member was available, the interviewers were instructed to interview a nursing home staff member who was involved in the residents’ daily care (staff proxy).

The MSHO evaluation included an interview of 2116 nursing home residents in Minnesota. Among them, 736 were enrolled in the MSHO program, 673 were nursing home residents in the same geographic area who were dually eligible for Medicare and Medicaid coverage but did not enroll in the MSHO program, and 707 were nursing home residents who were dually eligible for Medicare and Medicaid coverage in another part of the state but who did not have the opportunity to enroll in the MSHO program. All MSHO enrollees who volunteered to participate in the evaluation were included in the study sample.

The two control groups were selected based on the best match on several major characteristics with the MSHO sample (age, sex, race, previous nursing home residence since the start of MSHO, and their Medicare managed care enrollment status before MSHO enrollment). Similar to the EverCare evaluation, trained interviewers interviewed the residents. Whenever possible, the residents were interviewed. If the residents were too frail to respond, the interviewers were instructed to interview a significant family member who had regular contact with the residents. This
significant family member was selected and confirmed with the resident (family proxy). If no family proxy was available, the interviewers were instructed to interview a nursing home staff member who was involved in the residents’ daily care (staff proxy).

The same questionnaire was used in both evaluations. A component of the questionnaire measured the dependency on four ADL items: dressing, toilet use, transfer, and feeding. For each item, the response metric was: 1) resident did the task without help and without difficulty, 2) did the task without help but with difficulty, 3) required a little help from someone, 4) required a lot of help from someone, and 5) could not do the task at all. The ADL metric used in the MDS are: 0) independent, 1) supervision, 2) limited assistance, 3) extensive assistance, 4) total dependence, and 8) activity did not occur during the entire 7 days. Although these two metrics were not exactly the same, they were comparable. For our data analysis, we recoded the five-level response metric used in the interview and in the MDS into a three-level response metric. Figure 1 shows the crosswalk that we used in the recoding. We used this three-level coding for all analyses reported here.

For each nursing home resident, we obtained the MDS record most proximate to the time of the interview. For the MSHO interview, we did not record the interview date in the data file. However, because most of the MSHO interviews were conducted within the 2 months between early March and early May 1999, we assigned March 31, 1999 as the MSHO survey interview date for all MSHO respondents.

We used Kappa statistics to measure the agreement between the MDS and interview data collected from the EverCare and MSHO evaluations. We used multinomial logit regression analysis for risk adjustment to separate the effects of other covariates, such as resident demographics, from the effects of proxy responses. The dependent variable was a categorical variable with three possible outcomes: (a) the MDS rating of ADL difficulty was lower than the rating from interviews (MDS < interview), (b) the MDS and interview data had comparable ratings of ADL difficulty (MDS = interview), and (c) the MDS rating of ADL difficulty was higher than the interview rating (MDS > interview). We assigned the second group (MDS = interview) as the reference group in the analysis.

We included demographic variables (age, sex, and race), study group status (experimental and the two control groups), number of days between the resident interview and the MDS assessment (days off), and the additive ADL scores (dressing, toilet use, transfer, and feeding) from the MDS as independent variables for risk adjustment. The experimental group (EverCare/MSHO enrollees) served as the reference for “study group status,” and the number of days between the MDS assessment and interview more than 30 was the reference for “days off.” We did not include the cognitive functioning of nursing home residents in our risk adjustment because it is highly associated with the proxy status. We used odds ratios and 95% confidence intervals to report the degree of disagreement.

RESULTS

The final sample for this study was 3385 nursing home residents, including 1269 respondents in the EverCare evaluation and 2116 in the MSHO evaluation. Among them, 1200 interviews were conducted with the residents (group 1), 1070 interviews were conducted with family proxies (group 2), and 1115 interviews were conducted with staff proxies (group 3). Thirty-five percent were enrolled in either the EverCare or MSHO programs (experimental group), 32% were not enrolled in the EverCare or MSHO facilities (control-in group), and 33% were not enrolled in the non-EverCare or non-MSHO facilities (control-out group). Respondents who were in group 1 (resident interviews) were younger, more likely to be people of color and male,
and had a better ADL function compared with those in other two groups. Matched MDS assessments were distributed almost equally between periods before and after the interviews. Approximately 69% of the matched MDS assessments occurred within 30 days before or after the interviews and 96% within 45 days before or after the interviews.

Table 1 shows how the patterns of agreement and disagreement in ADL assessments between interview data and the MDS differed across the three data sources. We present the results as the percentage of cases in each of the respondent groups. The percentage of agreement in ADL assessments between MDS and interview data (MDS = interview data) was between 51% (dressing) and 82% (feeding) for data from resident interviews (group 1), between 66% (feeding) and 79% (dressing) for data from family proxy interviews (group 2), and between 62% (feeding) and 77% (dressing) for data from staff proxy interviews (group 3).

Data from resident interviews were less likely to have the same ADL assessment as the MDS as those in the three of the four ADL items (dressing, toilet use, and transfer) compared with the other two groups. For feeding, assessment data from resident interviews had a higher percentage of agreement with assessment in the MDS than those in the other two proxy groups.

The disagreement on ADL assessments between the MDS and interview data showed somewhat different patterns across ADL items. The MDS rated a lower level of disability (MDS < survey) in all four ADL items in group 3 (MDS vs staff proxies) than in group 1 (MDS vs residents) and group 2 (MDS vs family proxies). Conversely, the MDS rated a higher level of disability (MDS > survey) in three of the four ADL items (dressing, toilet use, and transfer) in group 1 (MDS vs residents) than in the other two proxy groups.

Table 1 also shows how the patterns of agreement and disagreement on the sum of the ADL scores between the interview data and MDS differed across the three groups. In general, the level of agreement was lower with the summed ADL than for any of the four individual ADL items. The MDS and interview data were less likely to have comparable ADL ratings in group 1 than the other two groups. Similar to the findings for the individual ADL items noted previously, the MDS was more likely to rate fewer ADL difficulties than interview data in group 3 than the other two groups. The MDS was more likely to rate additional ADL difficulties than interview data in group 1 than the other two groups.

The last three columns of Table 1 show the Kappa statistics for the four ADL items and the summed ADL. The Kappa statistics for the four ADL items were between 0.29 (group 1, feeding) and 0.52 (group 2, transfer), reflecting only fair to moderate agreement on ADL assessments between interview data and the MDS. Compared with group 1, groups 2 and 3 had higher Kappa statistics in three of the four ADL items (dressing, transfer, and feeding) and comparable Kappa statistics for toilet use. The Kappa statistics for summed ADLs were between 0.25 and 0.31, indicating a lower level of agreement between interview data and the MDS than for individual ADL items.

The level of agreement on ADL assessments between the MDS and interview across the three groups may be affected by other factors such as resident demographics, group status (control vs experimental), and level of disability. We used multinomial logit regression to isolate the effect of these variables from the effect of proxy assessments. Table 2 shows the results of the multinomial logit regressions. The top panel of Table 2 shows the odds ratios and 95% confidence intervals of having more ADL dependencies than apparent with the MDS (MDS < interview). The lower panel shows the odds ratios and 95% confidence intervals of having fewer ADL dependencies than evident on the MDS (MDS > interview). The reference group for these two panels was comparable ADL assessments between MDS and interview data (MDS = interview). The results were consistent with the findings reported in Table 1. After adjusting for the effects of other covariates in multinomial logit, the MDS was more likely to have fewer ADL difficulties in groups 2 and 3 than in group 1 and was less likely to have more ADL difficulties in groups 2 and 3 than in group 1.

**Discussion**

In this study, we investigated the effect of proxy responses on the accuracy of ADL assessments in the MDS. We matched interview data from two evaluation studies with MDS data to study the patterns of agreement and disagreement on ADL assessments between interview data and the MDS across three data sources. The overall agreement on ADL assessments between MDS and interview data was low to moderate, regardless of the data sources. The ADL assessments in the MDS were more often comparable with ADL assessments made by staff or family proxies than those made by residents’ themselves. The MDS was more likely to have fewer ADL difficulties than interview data obtained.
from staff proxies. The MDS was more likely to have more ADL difficulties than interview data derived from residents. These findings hold even after adjustment for other confounding factors using multinomial logit regression.

Findings from this study are consistent with findings from an earlier study on the effect of proxy responses in functional assessment data in the MDS. The report of ADLs in the MDS is compared with staff proxy interviews, considerable disagreement still occurs.

We tested the effect of an alternative recategorization of ADL items on our results to determine how our findings were affected by the way we recategorized the ADL items. The most confused level of ADL difficulty was “limited assistance” in the MDS assessment and “with a little help from someone” in the interview. This blurring may have resulted from issues surrounding how to handle “supervision” in the MDS. In this study, we recategorized supervision as independent. We tested an alternative recategorization by putting supervision into the limited assistance category. The extent of the discrepancy between the MDS and interview assessment was not accounted for by the biases. Therefore, error in functional assessment in MDS is a concern. The findings in Table 1 suggest that even when the MDS is compared with staff proxy interviews, considerable disagreement still occurs.

The discrepancy in the MDS and interview assessments between proxies and residents may reflect both error and bias. We tested the possible sources of bias using multinomial logit regression. Results of multinomial logit regressions showed that ADL assessments by family and staff proxies differed systematically from the assessments by the residents. These results suggest bias. However, the model fits (pseudo-$R^2$) of the multinomial logit were low, indicating that a large portion of the discrepancy in the MDS and interview assessments was not accounted for by the biases. Therefore, error in functional assessment in MDS is a concern. The findings in Table 1 suggest that even when the MDS is compared with staff proxy interviews, considerable disagreement still occurs.

We tested the effect of an alternative recategorization of ADL items on our results to determine how our findings were affected by the way we recategorized the ADL items. The most confused level of ADL difficulty was “limited assistance” in the MDS assessment and “with a little help from someone” in the interview. This blurring may have resulted from issues surrounding how to handle “supervision” in the MDS. In this study, we recategorized supervision as independent. We tested an alternative recategorization by putting supervision into the limited assistance category. However, this alternative recategorization generated lower concordance.

The extent of the discrepancy between the MDS and interview data raises concerns about the quality of the functional assessment data in the MDS. The report of ADLs should be more objective than other elements in the MDS.
such as pain, depression, and social activity, which require greater levels of subjective judgment by the observer. At the very least, these findings suggest a need for improved training and better communication on the MDS assessment mechanism to achieve higher levels of reliability. More thought should be given to moving from a rating system based on observations and inferences to one that requires more direct input from the residents.

ACKNOWLEDGMENT

Address correspondence to Terry Y. Lum, PhD, 105 Peters Hall, 1404 Gottner Avenue, University of Minnesota School of Social Work, St. Paul, MN 55108. E-mail: tlum@umn.edu

REFERENCES


Received November 18, 2003
Accepted January 27, 2004
Decision Editor: John E. Morley, MB, BCh

Downloaded from https://academic.oup.com/biomedgerontology/article-abstract/60/5/654/561422
by guest
on 22 June 2018