ORIGINAL ARTICLE

A review of evidence on the reliability and validity of Minimum Data Set data

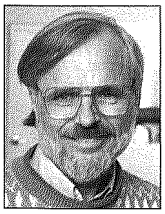
by J.W. Poss, N.M. Jutan, J.P. Hirdes, B.E. Fries, J.N. Morris, G.F. Teare, and K. Reidel

Jeff Poss, PhD, is an Assistant Research Professor in the Department of Health Studies and Gerontology, University of Waterloo. His research interests include older adults in community and residential settings, and the use of standardized assessments to better understand them.

Norma Jutan, MSc, is a PhD candidate in the Department of Health Studies and Gerontology, University of Waterloo. As a student in the Collaborative PhD Program in Aging, Health and Well-being, her research focuses on geriatric assessment with a particular emphasis on community-dwelling older adults.

John Hirdes, PhD, is a Professor in the Department of Health Studies and Gerontology, University of Waterloo and the Scientific Director of the Homewood Research Institute. He holds the Home Care Research and Knowledge Exchange Chair for Ontario.

Brant E. Fries, PhD, is a Professor of Health Management and Policy and Research Professor at the Institute of Gerontology, University of Michigan, and Chief of Health Systems Research for the Geriatric Research, Education, and Clinical Center at the Ann Arbor Veterans Affairs Medical Center.



John N. Morris, PhD, is the Co-Director of the Institute for Aging Research at Hebrew Senior Life, Boston, and Director of Social and Health Policy Research. He also holds the Alfred A. and Gilda Slifka Chair in Social Gerontological Research, the first faculty chair in social gerontological research in the United States.

Gary Teare, PhD, is the Director of Quality and Measurement and Analysis at the Saskatchewan Health Quality Council. His background includes leading the Ontario Hospital Report series for Complex Continuing Care, quality improvement in incontinence care in that sector, and research into factors related to best practice implementation in long-term care homes.

Kristen Reidel, BSc, is a research assistant at the Clinical and Health Informatics Research Group at McGill University where she is currently pursuing a Master of Science in Epidemiology. She completed a Bachelor of Science in Honours Health Studies at the University of Waterloo.

Abstract

This paper reviews the reliability and validity of the Minimum Data Set (MDS) assessment, which is being used increasingly in Canadian nursing homes and continuing care facilities. The central issues that surround the development and implementation of a standardized assessment such as the MDS are presented, including implications for health care managers in how to approach data quality concerns. With other sectors such as home care and inpatient psychiatry using MDS for national reporting, these issues have importance in and beyond residential care management.

Résumé

Le présent article analyse la fiabilité et la fiabilité de l'évaluation sur l'ensemble minimal de données (EMS), utilisée de plus en plus dans les centres d'hébergement et de soins de longue durée canadiens. Les principales questions qui entourent la création et l'adoption d'une évaluation normalisée comme l'EMD sont présentées, y compris les répercussions pour les gestionnaires de la santé quant à la qualité des données. Dans d'autres secteurs comme les soins à domicile et les services psychiatriques aux patients hospitalisés qui font appel à l'évaluation sur l'EMD pour les déclarations nationales, ces questions ont une importance qui dépassent ceux de la gestion des soins résidentiels.

The need for a uniform system of resident assessment in nursing facilities led to the development of a MDS in the United States starting in the late 1980s. The MDS was conceived as a standardized assessment instrument that would describe the important domains of health and care at an individual resident level, using the fewest data items possible. The MDS collects information on cognition, communication, vision, hearing, mood, behaviour, psychosocial, physical function, diseases, continence, health conditions, nutrition, dental, skin, activities, medications, and treatments and procedures, using about 400 data items.[1] Frontline clinical staff use the MDS to assess virtually all residents in U.S. nursing homes, with over 15 million assessments completed every year. As such, the MDS broke new ground in instrument design, deployment, day-to-day use and monitoring. Throughout the 1990s, as the MDS was rolled out and refined, a variety of research reports and related discussion appeared in the literature raising concerns about the quality of these data, fueling a debate that continues today. What evidence is there that this widespread implementation is likely to yield data in which clinicians and health executives can be confident? This paper reviews the published literature dealing with the central issues around data quality of the MDS.

The term itself can be confusing. While MDS represents a generic label for any measurement system designed with a minimum number of items, in the U.S. the widespread use of the nursing facility MDS instrument has led to its synonymous association. The developers suggested that the system of assessment and care planning guidelines be termed the Resident Assessment Instrument (RAI) and reserved the name "MDS" for the assessment instrument itself; others use these terms synonymously. Further, the international group interRAI, a research collaborative from about 30 countries,

has developed a family of assessment systems that includes the nursing facility instrument and operationally similar versions for home care, inpatient psychiatry, acute care and other settings. For this paper, the term MDS is used to mean the Minimum Data Set Instrument used in nursing homes, variously also known around the world as skilled nursing facilities, long-term care homes and complex continuing care hospitals/units.

With the MDS developed and initially implemented in the U.S., much of the work around its data quality also was done there. However, evidence from international use of the MDS has also contributed considerably to this literature. In Canada, mandated use of the MDS began in 1996 for all Ontario complex continuing care hospitals/units. Subsequently, it was chosen as the national reporting instrument for continuing care by the Canadian Institute for Health Information (CIHI), where it was slightly modified for application in Canada.[2] Its use in Canada is becoming widespread, as the majority of the provinces and territories have, or are planning, mandated implementations in their nursing home/continuing care sectors.

The MDS is an assessment with multiple applications, including care planning and outcome measurement at the individual level, as well as quality management and case mix/payment system classification at the organizational level. Data can be accumulated to the regional, provincial or national level, and assessment items from multiple domains allow stratified or adjusted comparisons on an "apples to apples" basis.

Quality measurement clinical benchmarking, public reporting and case mix classification using MDS data can be expected increasingly to be important in evidence-based health care management. Hence, the quality of the underlying data is a fundamental concern. The following points illustrate why managers and health executives should be concerned with the reliability and validity of data:

- the MDS can replace some of a facility's legacy clinical assessments, but without confidence in the data, clinicians will be hesitant to give up their traditional tools, resulting in redundancy and wasted effort;
- facilities being compared on quality indicators need to be confident in the accuracy of their own data and of those against which they are benchmarked;
- allocating funding resources based on MDS case-mix measures demands good quality data, the absence of which is at best random allocation and at worst allocation that might reward poor data quality; and
- considerable resources are required to support MDS assessment, including time spent on assessments as well as training and ongoing education of assessors, software and information technology support costs, and management oversight. This commitment demands acceptable quality data.

In addition, accountability agreements between regional or provincial funding bodies and providers will look to MDS data as a measurement source, for example, for program evaluation or policy development, but only if the data are trusted as fairly representing the things they purport to measure.

Measurement quality

Reliability represents the consistency of a measurement in its representation of a true value. While a number of methods can be used to estimate reliability, two common approaches are widely used in MDS studies. One involves inter-rater studies in which two people fully independently assess the same individual within a short period of time, and levels of agreement are calculated (adjusting for chance agreement). Studies that employ simultaneous assessors, rely on case-study-based evaluations or permit assessors to confer with each other will substantially inflate inter-rater reliability estimates. Inter-rater reliability must be conducted independently without communication between assessors in order to mimic real-life situations and to establish the confidence level clinicians and administrators can have in the assessments done by others at different times or from different organizations.

Another widely used method applies a measure of internal consistency (correlation of items that represent a common underlying concept), using the Cronbach's alpha statistic. This is appropriate for groups of parallel measures of a common underlying concept, such as different Activities of Daily Living (ADL) items. When items that are expected to group together are found to do so, this suggests that they are a reliable measure of a true underlying concept. Measures of internal consistency are useful because they provide a convenient statistical approach to evaluate data already available, without having to incur the significant costs or assessment burden associated with completing an inter-rater reliability study. However, this approach is limited by its applicability when the data include parallel-form measures.

Validity deals with whether an item truly measures the concept that it attempts to measure. While non-empirical tests of validity (face, content) have been applied to the MDS, published studies have typically dealt with the following types of validity: criterion, convergent and predictive. Criterion validity assesses the agreement between an instrument measure and an accepted "gold standard" measure, for example, the MDS cognitive performance scale and the Mini Mental State Exam (MMSE). Convergent validity measures associations that are expected, for example, indicators of cognitive impairment and a diagnosis of Alzheimer's disease. Predictive validity relates to a measure being associated with a future occurrence, for example, indicators of health instability being predictive of death.

Indicators of reliability and validity are both crucial to instilling confidence in a measurement system. Reliability is fundamental to instrument development, since a measurement will never be considered valid if it cannot be measured reliably. Alternatively, a measurement that has high

demonstrated reliability but little validity is of no practical use.

Problems with data quality in a multiple-domain instrument like the MDS may be attributed to two broad areas: the instrument and the assessor. During instrument development, reliability and validity testing identify problematic items that might be improved by re-wording, altering response choices or improved documentation and training; all give clearer direction and reduce subjective interpretation. An item that cannot be made reliable should be removed, unless it serves another purpose (e.g., encouraging a conversation with the resident). When an instrument yields acceptable data quality during development but not when more widely deployed in the field, improved training of assessors and vigilance by managers is needed. Indeed, any effort to implement a sophisticated system like the MDS should include a commitment to ongoing education, to support the continuing improvement of data quality and appropriate application of the instrument at all levels of the organization.

Reliability

Quality of measurement was central in the development of the MDS. The initial trial of the draft instrument included an inter-rater reliability study using 140 assessment pairs.[3] Of the 255 data items that could be examined for reliability, 55% showed acceptable to good performance. All scaled ADL items performed notably well, while other domain areas showed more mixed performance. Clinical feedback resulted in significant revisions, with about 20% of items being dropped and another 40% revised.

A later U.S. reliability trial,[4] also in the early 1990s, used a revised version of the MDS. About 123 rater pairs found acceptable or higher reliabilities for almost 90% of the items. ADL measures again demonstrated excellent reliability and only indicators of delirium were notably weak. Average reliabilities improved more than 50% from the earlier trial, a result the authors attribute to the exclusion or modification of weak items previously identified. The accumulated evidence from this and earlier work with previous forms of the MDS resulted in the MDS Version 2.0 (MDS 2.0) instrument, released in 1995.

In a follow-up inter-rater reliability study using the MDS 2.0,[5] new/revised items were found almost always to achieve acceptable to high reliability, and older unchanged items also were improved compared to previous studies, due largely to updated training materials.

While assessment items may show acceptable reliability among residents as a whole, they may show lower reliabilities among some sub-groups. This was examined for residents with significant cognitive impairment in an important study by Phillips and colleagues.[6] The study collected pairs of inter-rater reliability data on 147 residents for whom both assessors agreed on cognition level. MMSE scores were available to confirm further the designation of each case as either cognitively intact or impaired. Overall, a statistically higher level of inter-rater disagreement was found for the impaired residents when assessing functional status, communication and drugs/restraints. Mood/behaviour items did not demonstrate this effect. Items with subjective components, known to have poorer reliabilities, were not as affected by cognitive impairment as were stronger items, such as ADLs. The authors speculate that cognitively impaired residents may naturally exhibit more true variability (e.g., "sundowning"), making standardized instruments inherently less reliable in some populations. It is worth noting that when interRAI develops new instruments for distinct populations (such as inpatient psychiatry), further inter-rater reliability testing is conducted. It is not presumed that an item found to be reliable in one population will necessarily be reliable in a new one or that assessment practice patterns are equally reliable across sectors.

Further international evidence of the reliability of the MDS came from a seven-country study[7] that presented findings from the U.S., Japan, Denmark, Iceland, Italy, Sweden, and Switzerland. Results ranged from Switzerland, where 96% of items showed adequate or better reliability, to Japan and Sweden, with about 80% of items deemed reliable.

Inter-rater reliability for specific MDS domains or applications has also been measured, including case mix,[8-10] quality indicators[11] and incontinence[12]; all used accepted protocols and Kappa statistics and demonstrated strong reliabilities. The notable exception is the report by the U.S. Office of the Inspector General,[9] in which MDS assessments that were constructed based on the clinical record alone were compared for an exact match to a facility's MDS ADL measures and case-mix classification group assignment. Agreement rates under these conditions were low: 24% to 37%, respectively. However, the inspectors found no evidence that systematic gaming (false coding of items to achieve greater reimbursement) was occurring, with the rate of undercoding cost to matching that of overcoding, with no resulting significant difference in reimbursement.

Based on these earliest reliability findings, questions were raised as to the suitability of the MDS for research purposes.[13] Compared to rigorous and highly controlled research approaches, Teresi and Holmes[14] argued that MDS measures designed for clinical utility, in the absence of published validity studies, should make researchers extremely cautious. Representing the developers, Hawes and colleagues[15] acknowledged that researchers should always be concerned about data quality, but that no such distinct gulf exists between research and clinical data quality. They argued that the potential benefits of population level data provided by the MDS compel it to be used for important clinical and policy questions by researchers. Phillips and Morris[16] compared routinely collected MDS data from three states to data collected by trained research staff, and found measures of internal consistency for ADL and cognition to be similar. Further, they found ADL reliability measures to be similar for cognitively intact and severely cognitively impaired residents.

In the largest inter-rater reliability study ever conducted, with 5,700 U.S. nursing home residents from over 200 facilities, the research focus was not solely on reliability of

individual items, but also on composite quality indicators. This study was particularly important in the context of the use in the U.S. of quality indicators for public facility comparison. Of the 100 items used in the quality indicator construction, only 3 (difference in ADL comparing mornings to evenings, a do not hospitalize directive and diagnosis of paraplegia) were found to have weak reliabilities. Of the 12 constructed quality indicators, one was found to perform poorly (little or no activities) and one was marginal (infections).

Validity

As mentioned above, validity is a measurement property more challenging to demonstrate than reliability, especially with an instrument as broad as the MDS. With a variety of clinical domains, demonstrated validity in one domain does not automatically extend to others, and similarly, a single domain with weak validity does not necessarily reflect on the entire instrument.

Scale development for the important health domains in the MDS involved the validation of MDS items with gold standard measures. These criterion validity-based studies included the cognitive performance scale,[17,18] depression rating scale,[19] ADL hierarchy scale,[20] index of social engagement[21] and pain scale.[22] The MDS Changes in Health, End-stage disease and Symptoms and Signs (CHESS) scale is an example of a measure that shows excellent predictive validity, whereby greater health instability is predictive of death.[23] Further, an early criterion validity study, using the forerunner to the MDS 2.0 (the MDS+), compared gold standard measures of physical function, cognition, communication, mood and behaviour to equivalent indicators in the MDS, with only mood and behaviour failing to show strong evidence of validity.[24]

Two companion papers [25,26] applied a confirmatory factor analysis technique to early MDS data drawn from a single facility. This technique tests hypotheses regarding which items of the MDS belong together, and this study examined cognition, ADL, time use, social quality, depression and behaviour. Results found evidence of validity for all domains except social quality; however, poorer results were found among those cognitively impaired. For most MDS domains, it was a useful source of research data. This study included an inter-rater reliability component and found generally high reliabilities.

In another criterion validity study, Snowden and colleagues[27] found that measures of cognition, behaviour and ADL correlated well with gold standard measures using 140 subjects. This study also examined the ability of the MDS to detect change over time compared to gold standard measures and found the MDS cognitive performance scale to be more responsive than MMSE score, a six-item MDS ADL scale to be less responsive than the Dementia Rating Scale and both measures of behaviour to be generally unresponsive.

A large study, but one that used data from a pre-MDS 2.0 version, involved validation of the cognition measures of the MDS against MMSE scores.[28] Unlike some earlier derivation studies, this work did not use trained research nurse assessors, but used the facility-administered MDS data. While results were somewhat weaker than in the derivation work, they nonetheless correlated well.

Studies have examined whether the MDS is more likely to over-report (e.g., falls[29]) or under-report (e.g., urinary tract infections[30]) important health conditions, although the use of clinical records or other administrative sources as gold standards introduces additional measurement issues. Other specific clinical areas that have been examined in the literature include mixed validity evidence for nutritional status[31,32] and questionable construct validity for pressure ulcers.[33]

Depression measures in the MDS have received attention in validity studies previously described with generally acceptable findings.[24-26] Some researchers have found either individual mood items or the imbedded depression rating scale to be only weakly associated with other indicators of depression.[34-37] In a recent paper by Koehler and colleagues[38] using two versions of the geriatric depression scale compared to items in the MDS, they found the measures to be describing different dimensions of depression, suggesting a high degree of complexity in operationalizing depression in this population. Notably, sensitivity to depression indicators may differ depending on whether the measures are self-reported or by rater observation, something that is further complicated by high rates of cognitive impairment in these settings.

Pain is another important measure that has received considerable attention in the research literature, with some of the same issues (i.e., self-report measures and cognitive impairment). Engle and colleagues,[39] employing a non-representative sample and using the MDS as the gold standard measure of pain, found that nurses and aides tended to underreport pain, regardless of a number of resident characteristics, including cognition. In a small study of cognitively impaired residents, a proxy measure for pain (not validated as a gold standard measure) was found to be more sensitive and more highly correlated with analgesic use than were MDS pain items.[40] Cohen-Mansfield reports in two studies[41,42] that the MDS pain indicators associate well with other measures for those with little or moderate cognitive impairment, but may underestimate pain for those with severe cognitive impairment. Further, in a large sample of nursing home residents, research staff conducted a four-item pain interview and found that MDS-reported pain rates were much lower overall, especially for the most cognitively impaired.[43] However, as one of these researchers notes: "Given that there is no gold standard for the assessment of pain in severely cognitively impaired persons, the results with the severely impaired may only highlight the difficulty in pain assessment in this population in general, rather than reflecting on the validity of the MDS for this assessment."[41] Ascertainment bias (where measured values are systematically influenced by the skill or knowledge of the individual assessor) is a threat to validity and has been suggested for both pain and depression items of the MDS. Schnelle and colleagues[44] used research staff to investigate rates of a depression quality indicator in

two facilities, one closely associated with a university from which mental health services were contracted that showed a high depression rate, and a second with no such association that had a low rate. Additional assessment by research staff showed depression to not differ significantly between the facilities, suggesting that the additional expertise of the university consultants conditioned the setting for greater sensitivity to MDS items on depressed mood. It should be noted that the QI definition used in this study is no longer relied upon in the U.S. and has been replaced by a new version that examines worsening of mood or anxiety, and is further risk-adjusted to make comparison of sites more equitable. Similarly, pain measurement was examined for a broad sample of MDS-assessed residents, and after controlling for several socio-demographic and clinical factors, was found more likely to be reported in facilities with hospice provision,[45] suggesting ascertainment bias. However, the authors acknowledge that their results cannot prove this directly and that more study and sophisticated analyses are required.

Other forms of interRAI instruments have also been shown to have good reliability and validity as part of instrument development, for example, in home care,[46] post-acute care,[47] palliative care[48] and inpatient psychiatry.[49] These studies underline the enormous utility of equivalent validated measures that can be used across different settings.

Discussion

An important consideration with the MDS is the fact that each of the over 400 data items, alone or in clusters, can differentially show indicators of reliability or validity. Some types of measures are inherently stronger (e.g., observable, physical activities requiring help such as ADLs), with others weaker (e.g., ratings requiring measurement skill and judgment, such as pain in the cognitively impaired or highly unstable clinical conditions like fever). During development of an instrument, some items showing weaker data quality may be retained because they are seen as being of high clinical importance, such as was done with delirium items. Even for a specific domain, there will be a range of evidence, which must be balanced against the ambitious goals that the MDS was designed to meet. Scientific questions continue to be raised, for example, the editorial by Parmelee[50] that identifies issues with the use of MDS data for facility level quality improvement in the current U.S. regulatory environment. However, such healthy debates can only contribute to improvements in MDS data and Parmelee's call for better understanding of what works well and what does not is sound advice for researchers, frontline users and health executives. The multiple uses of MDS data, for quality measurement, case-mix measurement and payment, and clinical care planning, strongly encourage attention to data quality by multiple stakeholders. This paper presents evidence from the initial development of the MDS through the stable use of the MDS 2.0, a period spanning almost two decades. Ongoing development by interRAI has resulted in a revised suite of instruments that was announced in 2005, including a revised version of MDS known as the interRAI Long-Term Care Facility (interRAI LTCF) assessment instrument. Design considerations have included measures of reliability along with clinical utility and the capability to harmonize measures with instruments used in other settings. The interRAI LTCF is substantially shorter than the MDS and reliability performance was one of the central considerations in determining the inclusion or exclusion of items. Items that have historically provided weaker reliabilities, such as activity preferences, have been improved from their MDS 2.0 versions, and now show very good reliabilities. In addition, new items related to quality of life and subjective domains (e.g., personal goals of care) increase the emphasis on the perspective of the person being assessed. A forthcoming report on the multinational trial of the interRAI LTCF, as well as new instruments for home care, post-acute care, palliative care and mental health, demonstrates consistently strong cross-sector performance of core items that appear across care settings as well as specialized items unique to a specific sector. This type of work is critical for health executives committed to establishing a more integrated health system that supports effective cross-sector collaboration and continuity of care.

Data quality in the MDS will continue to reflect characteristics both of the instrument itself and of the assessors, their training and support. In Canada, CIHI contributes strongly to these latter characteristics, with documentation, training, support and data submission standards and checks. All stakeholders, including providers, regulators, researchers, payers and care recipients, have a shared interest in achieving assessment information that is of the highest quality possible. Consequently, ongoing education of clinical staff and health managers with respect to assessment practices and applications of the MDS is important. Health executives must ensure that all members of their organization share a commitment to obtaining high quality data to serve as a foundation of evidence to inform decision-making from the individual to the organizational level.

References

1. RAI MDS 2.0 and RAPs Canadian version: User's manual, second edition. Ottawa,ON: Canadian Institute for Health Information 2005.
2. www.cihi.ca
3. Morris JN, Hawes C, Fries BE, Phillips CD, Mor V, Katz S, et al. Designing the national resident assessment instrument for nursing homes. The Gerontologist 1990;30(3):293-307.
4. Hawes C, Morris JN, Phillips CD, Mor V, Fries BE, Nonemaker S. Reliability estimates for the minimum data set for nursing home resident assessment and care screening (MDS). The Gerontologist 1995;35(2):172-8.
5. Morris JN, Nonemaker S, Murphy K, Hawes C, Fries BE, Mor V, et al. A commitment to change: Revision of HCFA's RAI. Journal of the American Geriatric Society 1997;45(8):1011-6.
6. Phillips CD, Chu CW, Morris JN, Hawes C. Effects of cognitive impairment on the reliability of geriatric assessments in nursing homes. Journal of the American Geriatric Society 1993;41(2):136-142.

7. Sgadari A, Morris JN, Fries BE, Ljunggren G, Jónsson PV, DuPaquier JN, et al. Efforts to establish the reliability of the resident assessment instrument. Age and Ageing 1997;26 Suppl 2:27-30.
8. Ljunggren G, Fries BE, Winblad U. International validation and reliability of a patient classification system for long-term care. European Journal of Gerontology 1992;1:372-383.
9. Nursing home resident resource utilization groups. Washington, DC: Office of the Inspector General 2001.
10. Björkgren MA, Häkkinen U, Finne-Soveri UH, Fries BE. Validity and reliability of resource utilization groups (RUG-III) in Finnish long-term care facilities. Scandinavian Journal of Public Health 1999;27(3):228-234.
11. Mor V, Angelelli J, Jones R, Roy J, Moore T, Morris J. Inter-rater reliability of nursing home quality indicators in the U.S. BMC Health Services Research 2003;3:20.
12. Resnick NM, Brandeis GH, Baumann MM, Morris JN. Evaluating a national assessment strategy for urinary incontinence in nursing home residents: Reliability of the minimum data set and validity of the resident assessment protocol. Neurourology and Urodynamics 1996;15(6):583-98.
13. Ouslander JG. Maximizing the minimum data set. Journal of the American Geriatric Society 1994;42(11):1212-3.
14. Teresi JA, Holmes D. Should MDS data be used for research? The Gerontologist 1992;32(2):148-9.
15. Hawes C, Phillips CD, Mor V, Fries BE, Morris J. MDS data should be used for research. The Gerontologist 1992;32:563-4.
16. Phillips CD, Morris JN. The potential for using administrative and clinical data to analyze outcomes for the cognitively impaired: An assessment of the minimum data set for nursing homes. Alzheimer Disease and Associated Disorders 1997;11 Suppl 6:162-7.
17. Morris JN, Fries BE, Mehr DR, Hawes C, Phillips C, Mor V et al. MDS cognitive performance scale. Journal of Gerontology 1994;49(4):M174-82.
18. Hartmaier SL, Sloane PD, Guess HA, Koch GG, Mitchell CM, Phillips CD. Validation of the minimum data set cognitive performance scale: Agreement with the mini-mental state examination. Journals of Gerontology Series A: Biological Science and Medical Science 1995;50:M128-33.
19. Burrows AB, Morris JN, Simon SE, Hirdes JP, Phillips C. Development of a minimum data set-based depression rating scale for use in nursing homes. Age and Ageing 2000;29:165-72.
20. Morris JN, Fries BE, Morris SA. Scaling ADLs within the MDS. Journals of Gerontology Series A: Biologocal Science and Medical Science 1999;54(11):M546-53.
21. Mor V, Branco K, Fleishman J, Hawes C, Phillips C, Morris J, et al. The structure of social engagement among nursing home residents. Journals of Gerontology Series B: Psychological Science and Social Science 1995;50(1):P1-8.
22. Fries BE, Simon SE, Morris JN, Flodstrom C, Bookstein FL. Pain in U.S. nursing homes: Validating a pain scale for the minimum data set. The Gerontologist 2001; 41:173-9.
23. Hirdes JP, Frijters DH, Teare GF. The MDS-CHESS scale: A new measure to predict mortality in institutionalized older people. Journal of the American Geriatric Society 2003;51(1):96-100.
24. Frederiksen K, Tariot P, De Jonghe E. Minimum data set plus (MDS+) scores compared with scores from five rating scales. Journal of the American Geriatric Society 1996;44:305-9.
25. Casten R, Lawton MP, Parmelee PA, Kleban MH. Psychometric characteristics of the minimum data set I: Confirmatory factor analysis. Journal of the American Geriatric Society 1998;46(6):726-35.
26. 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 Geriatric Society 1998;46(6):736-44.
27. Snowden M, McCormick W, Russo J, Srebnik D, ,Comtois K, Bowen J, Teri L, et al. Validity and responsiveness of the minimum data set. Journal of the American Geriatric Society 1999;47(8):1000-4.
28. Gruber-Baldini AL, Zimmerman SI, Mortimore E, Magaziner J. The validity of the minimum data set in measuring the cognitive impairment of persons admitted to nursing homes. Journal of the American Geriatric Society 2000;48(12):1601-6.
29. Hill-Westmoreland EE, Gruber-Baldini AL. Falls documentation in nursing homes: Agreement between the minimum data set and chart abstractions of medical and nursing documentation. Journal of the American Geriatric Society 2005;53(2):268-73.
30. Stevenson KB, Moore JW, Sleeper B. Validity of the minimum data set in identifying urinary tract infections in residents of long-term care facilities. Journal of the American Geriatric Society 2004;52(5):707-11.
31. Blaum CS, O'Neill EF, Clements KM, Fries BE, Fiatarone MA. Validity of the minimum data set for assessing nutritional status in nursing home residents. American Journal of Clinical Nutrition 1997;66(4):787-94.
32. Bowman JJ, Keller HH. Assessing nutritional risk of long-term care residents. Canadian Journal of Dietetic Practice and Research 2005;66:155-161.
33. Zulkowski K. Construct validity of minimum data set items within the context of the braden conceptual schema. Ostomy/Wound Management 1998;44(10):36-40;42-4;46-7.
34. McCurren C, Dowe D, Rattle D, Looney S. Depression among nursing home elders: Testing an intervention strategy. Applied Nursing Research 1999;12(4):185-95.
35. McCurren C. Assessment for depression among nursing home elders: Evaluation of the MDS mood assessment. Geriatric Nursing 2002;23(2):103-8.
36. Anderson RL, Buckwalter KC, Buchanan RJ, Maas ML, Imhof SL. Validity and reliability of the Minimum Data Set Depression Rating Scale (MDSDRS) for older adults in nursing homes. Age and Ageing 2003;32:435-8.
37. Hendrix CC, Sakauye KM, Karabatsos G, Daigle D. The use of the minimum data set to identify depression in the elderly. Journal of the American Medical Directors Association 2003;4(6):308-12.
38. Koehler M, Rabinowitz T, Hirdes J, Stones M, Carpenter GI, Fries BE, et al. Measuring depression in nursing home residents with the MDS and GDS: An observational psychometric study. BMC Geriatrics 2005;5:1.
39. Engle VF, Graney MJ, Chan A. Accuracy and bias of licensed practical nurse and nursing assistant ratings of nursing home residents' pain. Journals of Gerontology Series A: Biological Science and Medical Science 2001;56:M405-11.

40. Fisher SE, Burgio LD, Thorn BE, Allen-Burge R, Gerstle J, Roth DL, et al. Pain assessment and management in cognitively impaired nursing home residents: Association of certified nursing assistant pain report, minimum data set pain report, and analgesic medication use. Journal of the American Geriatric Society 2002;50(11):152-6.
41. Cohen-Mansfield J. The adequacy of the minimum data set assessment of pain in cognitively impaired nursing home residents. Journal of Pain Symptom Management 2004;27:343-51.
42. Cohen-Mansfield J. Nursing staff members' assessments of pain in cognitively impaired nursing home residents. Pain Management Nursing 2005;6(2):68-75.
43. Chu L, Schnelle JF, Cadogan MP, Simmons SF. Using the minimum data set to select nursing home residents for interview about pain. Journal of the American Geriatric Society 2004;52(12):2057-61.
44. Schnelle JF, Wood S, Schnelle ER, Simmons SF. Measurement sensitivity and the minimum data set depression quality indicator. The Gerontologist 2001;41:401-5.
45. Wu N, Miller SC, Lapane K, Gozalo P. The problem of assessment bias when measuring the hospice effect on nursing home residents' pain. Journal of Pain Symptom Management 2003;26(5):998-1009.
46. Morris JN, Fries BE, Steel K, Ikegami N, Bernabei R, Carpenter GI, et al. Comprehensive clinical assessment in community setting: Applicability of the MDS-HC. Journal of the American Geriatric Society 1997;45(8):1017-24.
47. Morris JN, Jones RN, Fries BE, Hirdes JP. Convergent validity of minimum data set-based performance quality indicators in postacute care settings. American Journal of Medical Quality 2004;19(6):242-7.
48. Steel K, Ljunggren G, Topinkova E, Morris JN, Vitale C, Parzuchowski J, et al. The RAI-PC: An assessment instrument for palliative care in all settings. American Journal of Hospice and Palliative Medicine 2003;20(3):211-9.
49. Hirdes JP, Smith TF, Rabinowitz T, Yamauchi K, Perez E, Telegdi NC, et al. The Resident Assessment Instrument-Mental Health (RAI-MH): Inter-rater reliability and convergent validity. Journal of Behavioral Health Services Research 2002;29(4)419-32.
50. Parmelee PA. Quality improvement in nursing homes: The elephants in the room. Journal of the American Geriatrics Society 2004;52(12):2138-40.