Temporal and Geographic variation in the validity and internal consistency of the Nursing Home Resident Assessment Minimum Data Set 2.0

Vincent Mor, Ph.D.
Orna Intrator, Ph.D.
Mark Aaron Unruh, MA
Shubing Cai, Ph.D.

Department of Community Health and the Center for Gerontology & Health Care Research
Public Health Program, Alpert Medical School, Brown University

June, 2010
Minimum Data Set Consistency and Validity

ABSTRACT

Background. The Minimum Data Set (MDS) for nursing home resident assessment has been required in all U.S. nursing homes since 1991 and universally computerized since 1998. Initially intended to structure clinical care planning, uses of the MDS have expanded to include policy applications such as case-mix reimbursement and quality monitoring and research. The purpose of this paper is to summarize a series of analyses examining the internal consistency and validity of the MDS data as used in the “real world” in all US nursing homes between 1999 and 2007.

Methods. We used person level linked MDS and Medicare denominator and all institutional claims (part A) including inpatient (hospital and SNF) claims files for all Medicare fee-for-service beneficiaries residing in U.S. nursing homes during the period 1999 to 2007. We calculated the positive predictive value (PPV) of diagnoses taken from Medicare hospital claims and from the MDS among all new admissions from hospitals to nursing homes and then the internal consistency of items within the MDS that logically should be related. We also tested the internal consistency of commonly used MDS based multi-item scales and examined their predictive validity viz. one year survival. Finally, we examined the relative completeness of the MDS discharge record, specifically the extent to which it corresponds to the information on Medicare hospital claims.

Results. Each year there were some 800,000 new admissions directly from hospital to US nursing homes. The PPV of Medicare hospital diagnoses and MDS based diagnoses were between .6 and .7 for major diagnoses like CHF, hypertension, diabetes. Internal consistency of the ADL items on the MDS with other MDS items measuring impairments and symptoms exceeded .9. Multi-item scale alpha reliabilities stratified by ADL and Cognitive levels were around .6 or higher and the CHESS index, a summary measure of frailty was highly predictive of one year survival. The correspondence of the MDS discharge record indicating hospitalization and the existence of a Medicare hospital claim improved between 1999 and 2007, but over 20% of MDS discharges indicating hospitalization had no corresponding Medicare hospital claim.
Minimum Data Set Consistency and Validity

**Summary.** The MDS demonstrates a reasonable level of consistency both in terms of how well MDS diagnoses correspond to hospital discharge diagnoses and in terms of the internal consistency of functioning and behavioral items. The level of reliability and validity demonstrated by the scales suggest that the data can be useful for research and policy analysis, but to date the MDS discharge tracking record should probably not be used as a substitute for Medicare hospital claims data. With the introduction of the MDS 3.0 in the Fall of 2010, it will be important to monitor its performance with respect to consistency, reliability and validity using these results as a baseline that should be exceeded.
Minimum Data Set Consistency and Validity

Introduction

Originally developed in 1990 and mandated for nationwide use in response to the Nursing Home Reform Act of 1987, the Resident Assessment Instrument (RAI) Minimum Data Set (MDS) 2.0 has been in use since 1997 and computerized into a national repository since October of 1998.[1, 2] Initially designed as an instrument to summarize a detailed clinical assessment as a prelude to care planning, it was not long before its use was adapted to case-mix reimbursement as a means of setting daily payment rates in both the Medicare program as well as in nearly 40 state Medicaid programs.[3, 4] This was followed by the creation and public reporting of quality measures based upon facility aggregates of selected MDS data items.[5-7]

As a result of these expanded applications, the system of records governing the RAI became increasingly complicated. Under the original legislative mandate, the Centers for Medicare and Medicaid Services (CMS) required that a comprehensive resident assessment be completed within 14 days of admission to the facility. This assessment was to be completed at least quarterly thereafter meaning that admissions for short term stays would not have a documented assessment since it was not required.[1, 8] A discharge record was introduced around the time that computerization was mandated in 1998 and specialized Medicare assessments were introduced with the introduction of the Skilled Nursing Facility Prospective Payment System in order to document patients’ level of need for care and rehabilitation within a few days of admission and at regular intervals until discharge or on the standard schedule for long stay residents no longer covered by Medicare. [9]

The increased complexity along with the requirement for submitting computerized MDS records meant that even small facilities had to purchase computers and software or make arrangements for data entry of and transmission of their data to the central CMS repository. Larger facilities and multi-facility chains purchased or developed specialized software with a wide variety of capabilities, some automatically generating resident care plans so facilities were in compliance with regulations, others that automatically updated unchanged fields for
Minimum Data Set Consistency and Validity

quarterly assessments and still others that checked for data internal consistency.[10] The sophistication required to differentiate between types of records and to link admissions and continuing stay records to discharges is considerable, particularly since the definition of a discharge is ambiguous. For facilities that engaged in data based quality improvement efforts, there was the possibility for checking the accuracy and consistency of their data, but there is little evidence that this was a common practice.[11, 12]

Field testing of the MDS and the quality measures revealed generally good levels of inter-rater reliability among those facilities that agreed to participate in these large and demanding research studies.[13-15] However, numerous smaller studies exploring the validity of the data elements that make up some of the quality measures that are publicly reported by CMS have revealed substantial problems with the validity of the indicators and variability in how the data elements are recorded.[16, 17] At the same time, others using the MDS data for epidemiological and health services research studies have found that, in aggregate, the data behave as expected with respect to the creation of summary scales and the predictors of hospitalization and mortality.[18-22] The fact that the MDS data can perform as expected in broad scale research studies while not having a high degree of sensitivity in measuring quality is perfectly possible both because there is considerable inter-facility variation in data quality, even amongst facilities that volunteer, and because statistical significance in epidemiological studies can still substantially underestimate the strength of the relationship due to measurement error.[15, 23, 24]

In spite of ongoing questions about the validity of the MDS data and the related quality measures being publicly reported, CMS and the industry are committed to an MDS rooted in resident assessment but which is used for both payment and quality monitoring purposes. Indeed, a revised version of the MDS more focused on the “resident’s voice” was tested and refined for several years and scheduled for national implementation in October, 2010.[25, 26] This version requires staff to directly ask residents questions if they are able to respond while at the same time retaining many of the features of the MDS 2.0 that were used to calculate the case-mix measures necessary to determine facility reimbursement levels. While it is highly desirable to incorporate the
Minimum Data Set Consistency and Validity

resident directly into the assessment process, the proportion of residents able to respond directly to questions is likely to vary substantially across facilities, introducing yet another level of complexity into the interpretation of the data.[27]

Now that the MDS 2.0 is being replaced by the MDS 3.0, it is appropriate to consider selected aspects of the level of internal consistency achieved over the course of a decade of “real world” use along with the reliability and validity of some of the key scales and measures that have been constructed and widely used for reporting and research. The purpose of this paper is to examine the properties of the MDS 2.0 and how they’ve varied over time and geographically. As has been well documented by researchers working with Medicare claims data, all administrative data systems have weaknesses and each change in the regulations underlying their use is likely to alter some aspect of these data in meaningful ways. After over a decade of use, the MDS 2.0 has reached relative maturity. In several years, it will be useful to compare the performance of the MDS 3.0 on some of the parameters we present here in order to assess its development and better understand its strengths and weaknesses.

Methods

Overview. We used person level linked MDS and Medicare denominator and all institutional claims (part A) including inpatient (hospital and SNF) claims files for all Medicare fee-for-service beneficiaries residing in U.S. nursing homes during the period 2000 to 2007.¹ We compared diagnoses on the hospital claim with those on the MDS following admission from hospital. We used the Residential History File (RHF) methodology to compare the discharge locations specified on the MDS discharge record with that obtained using the RHF.[28] We also examined the rate of internal consistency of various MDS items that clinically should be highly consistent and finally examined the *alpha* reliability of commonly used MDS based scales and their association with subsequent mortality.

¹. Data were assembled under CMS Data Use Agreement #18900 associated with the NIA funded Program Project grant *Shaping Long Term Care in America* [AG #27296]
Minimum Data Set Consistency and Validity

*Sample.* We matched MDS assessments to the Medicare enrollment record using age, sex, date of birth, Health Insurance Claim number (Medicare #) and Social Security number when the Medicare# was not available for all MDS records associated with nursing home residents age 65 or over. We use an algorithm to identify unique individuals in the stream of MDS records by matching selected identifiers like the HIC# and the Social Security# along with sex and date of birth. The more MDS records a person has with comparable matching variables, the more confident we are that the records really pertain to a unique individual rather than representing a coding error. Thus, the longer the study period the more likely we have an identifier that matches to the Medicare enrollment record. We calculate match rates as the proportion of MDS identified individuals matched to the CMS denominator file containing the correct Medicare#, gender and date of birth. As can be seen in Table 1, match rates in all years (1999-2008) among those 65 and over exceed 95% with slightly higher numbers pertaining to earlier years, precisely because we have more data to correct “errors”.

After linking MDS and Medicare denominator files, we derived a RHF for all the residents in the cohort. The RHF was based on all Medicare inpatient hospital, Skilled Nursing Facility, out-patient, home health and hospice claims to these individuals, linked chronologically, and according to the order of data reliability, to form of a personal history. We excluded Medicare beneficiaries for a study (calendar) year if they had ever been a member of a Medicare Managed Care plan since those individuals’ utilization events are not captured in the standard Medicare fee for service claims systems. From this population of Medicare beneficiaries with at least one MDS record over the period 2000 to 2007, we developed two different analysis samples, each stratified by year. First, we separately linked hospitalization claims of individuals who had not had a prior MDS record, with the MDS record immediately following their discharge from the hospital. This allowed us to associate diagnoses listed in the Medicare hospital claim to the MDS admission assessment record. In addition to comparing diagnoses across the Medicare hospital claim and the MDS, we used the same admission MDS record to examine the internal consistency of selected MDS items and the inter-item consistency of MDS items that have been reported as representing multi-item scales. Secondly, we selected the population of individuals
Minimum Data Set Consistency and Validity

with an MDS admission and followed them over time to identify those with an MDS discharge and/or a
subsequent hospitalization in order to determine the rate at which MDS discharge records matched to a
hospitalization record, and whether a hospitalization record was preceded by a MDS discharge.

**Measures.** We present the results of three different sets of analyses using data from both the MDS
records and the Medicare claims. For the purpose of these analyses we view the Medicare claims as the “gold
standard” both for the diagnoses and the dates of service from inpatient and SNF claims. From Medicare claims
we use the ICD-9 diagnoses coded on the hospital claim which allows for up to 10 different discharge diagnoses
and an admission diagnosis. We rely heavily on the dates of admission and discharge from the hospital, since,
in determining the “validity” of the MDS discharge destination of hospital, we examine both the exact date
match of the MDS discharge date and the Medicare hospital claim admission date.

We compare the presence of selected diagnoses on the hospital discharge claim to those indicated on the
MDS admission record which uses a “check box” approach rather than ICD-9 coding. The groups of ICD-9
codes on the Medicare claim were contrasted with the presence of a positive “check box” on the MDS for an
appropriate diagnosis (e.g. heart failure on the MDS was equated to heart failure and cardiomegaly – ICD-9
codes 398.91, 402, 404, 428). In selecting which MDS “check boxes” to compare to hospital discharge
diagnoses, we focused on those which were least ambiguous with respect to the cross-walk with ICD codes and
built upon our earlier work in this area.[29]

Next, we tested the extent to which data on the MDS admission and discharge records correspond to the
sequence of Medicare claims data, including dates of re-hospitalization on the MDS discharge record and dates
of re-admission in the Medicare hospital claims. To the extent possible given the limitations of the Medicare
claims, we also cross-walked the reason for discharge code on the MDS discharge record and the sequence of
Medicare claims and/or the date of death in the Medicare denominator record. Since dates of service are
frequently one day off in light of how dates of discharge and admission are treated (can’t be billed for two

8
Minimum Data Set Consistency and Validity

inpatient services on the same day), we allowed a three day non-exact match tolerance in determining the rate of “exact correspondence” between MDS events and Medicare claims based events.

We examined the utility of the MDS discharge tracking form in identifying residents who died, and residents who were hospitalized, based on the reason for the discharge on the tracking form. We compared the MDS discharge destination with the location that we observed the resident at based on the RHF since the RHF provides daily information on location of care, based on all Medicare part A claims, Medicare denominator (eligibility file), and the dates in which MDS assessments were conducted and the reason for the assessment. Two comparisons were made: the first assumed that the RHF, replete with Medicare claims data, was the “truth”, and examined how well the MDS discharge assessments corresponded to these discharges. We next assumed that the MDS was correct and estimated the correspondence again. If one were to use only MDS in an analysis examining hospitalizations, it is critical to know whether and by how much estimates would over or underestimate hospitalization rates.

Finally, we examined the frequency with which an MDS is not filed during a Medicare stay, based upon the existence of a Medicare SNF claim as another means of assessing how much data are missing or incorrect (relative to Medicare claims) if on relies only upon the MDS records.

Based only upon data in the MDS admission assessment, we examined the internal consistency of selected items within the MDS that should logically correspond to one another. For example, residents who have no dependency in any Activities of Daily Living should not be bed-bound, hemiplegic or have no ability to move in their own bed independently. While probably a less perfect match, we tested those with a Cognitive Performance Score suggestive of severe impairment to determine if they are likely to have a diagnosis of dementia and/or Alzheimer’s disease.[18] We went somewhat further afield and compared those with aphasia with respect to their expressive communication abilities, those with arthritis with respect to join pain, those receiving a diuretic with respect to having been coded as having edema. For each of these internal consistency checks we calculated specificity, sensitivity and the PPV of each predicated relationship.
Minimum Data Set Consistency and Validity

We tested the internal consistency and alpha reliability of several multi-item scales that have been characterized in the literature based upon the MDS, using the same admission MDS. As was done in an early paper comparing the performance of some of these scales, we conducted the analyses stratified by median value of both the Activities of Daily Living (ADL) Scale and the Cognitive Performance Scale (CPS).[30] The rationale for calculating the internal consistency of these items separately within strata of ADL and CPS relates to the level of variation on measures such as depression, pain, social engagement, etc. which will be very different based upon the individuals’ basic physical and cognitive function.

Finally, we examined the predictive validity of several composite measures of functioning and frailty that have been developed including ADL, CPS and the CHESS.[31, 32] All three were based upon admission MDS data and were used to predict one year survival using the vital status information included in the Medicare denominator file regardless of whether the individuals remained in a nursing home or not. These analyses were stratified by age to test the independent effect of the frailty measures on one year survival, controlling for age.

Analytic Approach. We calculated the sensitivity, specificity and positive predictive value (PPV) of the MDS diagnoses relative to the “gold standard” of the Medicare hospital claims diagnosis for each year and separately for each state, allowing us to test both the effect of time and geography. The same approach was adopted in comparing the inter-item consistency. All analyses were performed using SAS version 9.2 and STATA 10.0.

Results

Table 2 presents a sample description of the new admissions entering US nursing homes in 2000, 2002, 2004 and 2006 based upon the Medicare denominator file information including date of birth, gender and race. As can be seen, the average age at first nursing home entry is some 81 years of age and this has remained fairly stable over the course of the decade. On the other hand, we are seeing a substantial increase in the number of diagnoses reported on the Medicare hospital claim preceding patients’ first admission to a nursing home.
Minimum Data Set Consistency and Validity

Interestingly, while the age distribution didn’t change much over the period (~ 20% for those 65-74 and over 30% for those over 85), the proportion of male admission did increase by over 2 percentage points.

Table 3 presents the results of comparing the MDS diagnoses on the admission assessment with those on the Medicare hospital claim discharge diagnoses for selected diagnoses. Also presented are the inter-quartile range and the 25th and 75th percentile reflecting variation across states in the average PPV. We conducted the analyses for all years between 2000 and 2006 but since the pattern of results was quite similar across all years, we are only presenting the most recent year. With a few exceptions, most of the diagnoses have PPV in excess of .6. Heart failure, diabetes and COPD/asthma/emphysema all had high PPV levels while depression, stroke and any dementia had much lower PPV levels. The PPV of Parkinson’s Disease is one which changed substantially over the study period, from .76 in 2000 to .60 in 2006, with a relatively high inter-quartile range but diabetes also declined over the period [data not shown; available from authors]. By and large, the inter-state variation as measured by the inter-quartile range for most of the comparisons is relatively small, particularly for those diagnoses with high levels of PPV.

To address the issue of inter-facility variation in the correspondence between Medicare hospital discharge diagnoses and MDS based admission diagnoses, we calculated the PPV at the level of the individual facility for facilities with a minimum of 100 admissions during the course of 2006. Figure 1 presents the distribution of facilities with various levels of PPV for the diagnosis of heart failure. As can be seen, most facilities with large numbers of admissions from hospital in the year have reasonably high levels of PPV when comparing Medicare hospital claims diagnoses with diagnoses on the admission MDS. Indeed, nearly 70% of facilities have a PPV in excess of 70%, reflecting the high average of .78 for heart failure in 2006.

Table 4 presents the results of the internal consistency analyses comparing how well MDS items that should be logically related actually do correspond in terms of the positive predictive value, along with the inter-state variation in the positive predictive value of the association between the two variables. Since we observed no consistent trend in the PPV values for these comparisons over time, we present only the data from 2006.
Minimum Data Set Consistency and Validity

although more detail is available from the authors. All the ADL related items demonstrate very high levels of internal consistency that has been very consistent over time. Interestingly, the correspondence between having a Cognitive Performance Scale score of zero (no obvious signs of cognitive impairment or memory loss) and the presence of a check-box diagnosis of Alzheimer’s disease or any dementia is quite high (>0.95) and the inter-quartile range based upon state level average PPV levels is very small, suggesting comparably high rates of association across the country. Communication, joint pain, vision, asthma/COPD/emphysema were only moderately associated with their respective criterion variables but cancer and chemotherapy and pressure ulcer care and pressure ulcers were moderately highly associated.

Researchers have constructed various multi-item scales from the item set in the MDS other than those measuring activities of daily living, mobility and cognitive functioning. Using data from the new admission cohort in 2007, we calculated the standardized alpha inter-item consistency coefficient for the social engagement scale, the mood (depression) scale, the behavior problem index and the pain scale.[19, 33, 34] Table 5 presents the results of our analyses, revealing standardized alpha coefficients relatively close to those reported by the scale developers stratified by the median ADL scale and CPS scores. The difference between the response patterns among the cohorts defined by the intersection of high and low ADL and CPS is not large, but it does reveal interesting patterns. As expected, consistent with the expectation that staff have more difficulty assessing cognitively impaired residents, the standardized alpha coefficient is consistently lower among those with low CPS. The pattern with respect to high vs. low ADL performance is more subtle; while the low ADL cohort reveals lower reliability for social engagement and depression, the difference is quite small for behavior and seems to move in the opposite direction for pain intensity.

Various “validity” studies of the MDS and its applications have been undertaken, often comparing “research” measurements done by clinicians or research assistants as compared to information in the most recent MDS in the residents’ files. Another approach to testing the construct validity of some aspect of the MDS is to examine the relationship between selected items and concepts which the literature and clinical
Minimum Data Set Consistency and Validity

practice tells us should be related to readily measured “objective” outcomes like death or hospitalization. We tested the “predictive validity” of the CHESS scale, designed as a composite measure of change of medical stability, frailty and clinical acuity, to predict mortality amongst frail elders in institutional settings.[31] We identified new admissions to US nursing homes in 2006 and monitored them for at least one year to determine whether they’d died according to the Medicare enrollment file. Figure 2 summarizes the relationship between quartiles of the CHESS scale and one year mortality, stratified by age upon admission. As can be seen, there is a doubling in the one year mortality rate among 85 year olds between the lowest and highest quartile of the CHESS scale, from around .30 to .60. Perhaps because this is a new admission cohort, many of whom do not remain nursing home residents but are re-hospitalized and die or return under hospice care, we see a strong monotonic effect of chronological age categories on one year mortality within quartiles of the CHESS score until the highest risk CHESS category is reached, at which point age doesn’t appear to matter. We conducted comparable analyses of the predictive value of the long form ADL scale as well as the Cognitive Performance Scale and both were found to be strongly related to one year mortality, although not as strong as the CHESS since it was designed precisely to be predictive of survival.

Figure 3 presents the proportion of Medicare paid hospitalizations of fee-for-service beneficiaries which were recorded by an MDS discharge tracking form, indicating discharge to a hospital, within 3 days of the inpatient admission day (either before or after). As can be seen, the rate of MDS reporting of a hospital discharge rose from about 81% in 1999 to almost 90% by 2007, with the largest improvement occurring around 2001.

However, when we examined all the MDS discharge tracking forms indicating discharge to hospital, we found that there were about 30% more MDS discharge tracking forms than hospitalizations indicated by claims in each year. In 1999 there were about 1.1 million MDS tracking forms indicating a hospitalization but 23% of them did not have an associated Medicare hospital claim. While the number of MDS discharge records indicating hospitalization increased to 1.4 million, the proportion of those without a hospital claim remained at
Minimum Data Set Consistency and Validity

about 22%. We further investigated where those persons who had the MDS discharge indicating hospitalization were when it was filed. Of those that had a hospital claim, about 75-79% were on the same day, and about 15-18% had a claim within 1-7 days after the MDS discharge. The remaining 6-10% had a discharge MDS that was filed during their hospital stay or even after the hospital stay ended. Of those MDS discharges which did not have a Medicare inpatient claim (~320 thousand in 2007), about 81-87% were in a nursing home, and a rising number (from about 11% in 1999 to over 17% in 2007) were in the ED either receiving emergency care, or having an observation stay. The proportion of MDS assessments that could be matched to inpatient claims varied across states in 2005 between 66.3% in Arizona to 85.6% in Kansas.

Discussion

We undertook a comprehensive, data based approach, to testing the consistency and utility of the MDS for administrative reporting and research uses. To do so, we used national Minimum Data Set Registry data merged with Medicare enrollment files and hospital and SNF claims data covering 1999 through 2007 to assess the “validity” of the MDS record sequencing and diagnostic information as well as the internal consistency and validity of the MDS items and the clinical research scales that have been developed. First, the match rates between MDS data and Medicare records exceeds 95% for the population 65 and over and there is very little inter-state variation in the match rate. Second, many of the “check box” diagnoses on the MDS correspond reasonably well (PPV > .7) against the gold standard of the Medicare hospital IDC-9 diagnoses, although without the obvious precision of an ICD code. Third, the data items within the MDS record that should be internally consistent (e.g. measures of physical functioning) are (PPV > .95) and those where there is less expectation of agreement have lower, but still reasonably high levels of agreement. Fourth, the internal consistency of proposed multi-item scales included in the MDS were found to be moderate and relatively constant across very different groups of patients with respect to cognitive and physical functioning. Fifth, we validated composite an acuity and frailty score, the CHESS scale, predictive of death and found it to be highly related to one year mortality based upon Medicare records, even after stratifying for age. Sixth, in examining
Minimum Data Set Consistency and Validity

the completeness and validity of the MDS Discharge Record information relative to Medicare records, we found that when Medicare records indicated a beneficiary had been hospitalized or entered a nursing home under the SNF benefit or died while in the nursing home, there tended to be a corresponding MDS Discharge Record and this improved over time. However, there were many more hospitalizations according to the MDS discharge tracking record than were substantiated by Medicare records. While there has been some improvement in the rate of completeness of discharge records over the last decade, there remains considerable inter-state difference. The paragraphs below discuss the implications of these results for the broad scale use of the MDS for payment, quality monitoring and research and for those charged with monitoring the implementation of the new MDS 3.0 being introduced in October of 2010.

Our finding that major diagnoses noted in the MDS are reasonably consistent with the diagnoses enumerated in the Medicare hospital claim replicates our finding from the early days of the MDS when only a few states were computerizing assessments.[29] The MDS manual stipulates that MDS diagnoses are those that affect treatment or function which is consistent with the instructions hospital coders adhere to under Medicare billing.[35] We did observe a large increase in the number of diagnoses listed on the hospital claim but it didn’t really influence the observed PPV’s which we found to be relatively stable over time. The increase in hospital diagnoses occurred presumably because of changes in reimbursement policies and the expanded use of quality measures which may have pushed hospitals’ coders to note increasingly specific diagnosis and procedure codes. Thus, our findings suggest that use of cardio-vascular disease, diabetes, Alzheimer’s disease and several others will yield research results consistent with use of Medicare claims data. While there is some inter-facility variation in the PPV between the two sources of diagnostic information, by and large it appears reasonable to use MDS diagnostic data to predict outcomes like hospitalization, discharge home or even death.

Almost since the initial design of the MDS as the summary of the resident assessment mandated by Congress as part of the Nursing Home Reform Act of 1987, disputes about the reliability, validity and accuracy of the data have raged.[2, 36-39] While a number of large reliability trials consistently found moderate to
Minimum Data Set Consistency and Validity

excellent levels of inter-rater reliability between research and staff nurses, others studies have found poor
correspondence between facility medical records, patient observations and the data in the MDS.[14, 15, 40, 41]
Some have noted substantial inter-state and inter-provider [42] variation in data quality and completeness and
that facilities which participated in reliability studies differed substantially from those that didn’t.[15, 23] Our
data do reveal some inter-state variation in agreement rates, but in general we find that data on items’ internal
consistency and agreement with hospital diagnoses are reasonably strong and consistent across states. As
importantly, the research scales on mood, behavior, social engagement and pain do seem to be based upon
consistent patterns of inter-relationships among the items since they were consistently observed across very
different clinical populations.

In light of our findings how should the MDS data be used? As noted, the MDS was designed to
document and guide a uniform resident assessment process for the purpose of developing a care plan.[1] As
such, the data should guide individual clinical decisions. Clearly the MDS doesn’t do that since the data are not
updated in real time, rather only once a quarter to represent a snapshot in time of the resident’s condition. The
MDS wasn’t supposed to replace the more dynamic medical record and nursing notes. When initially designed
over 20 years ago, the acuity and risk of change in clinical status was much lower than today’s more clinically
complex nursing home population. The question remains then – how good is this “snapshot” and what are the
implications of using it for reimbursement, monitoring providers’ quality performance or research?

Medicare and nearly 40 states’ Medicaid programs currently use MDS data to apply some form of case
mix reimbursement that increases payment rates as a function of the acuity and functional limitations of the
residents.[43-45] Medicare applies MDS to determine the exact payment to a given facility on behalf of a
specific patient while most state programs apply case-mix adjustment at the aggregated level of the facility.
Zinn and colleagues concluded that adopting this reimbursement model is associated with greater resident
acuity, suggesting improved access for sicker residents or more aggressive “up-coding”. Interestingly, the one
research audit done to address this issue was done by the General Accountability Office. They found as much
under as over coding of patients’ conditions relative to nurses notes or research staff assessments, a finding that is consistent with an analysis of the directionality of inter-rater reliability errors in the MDS of nursing staff relative to research nurses.[23] Thus, use of the MDS for reimbursement is not substantially different from using Medicare hospital claims for the application of prospective payment rates, with respect to overall accuracy.

Since the early part of the last decade, the Centers for Medicare and Medicaid Services has been using MDS data to create and publicly report quality measures at the level of the facility, contingent upon there being a sufficient number of residents in the home. Such aggregated measures can tolerate a certain level of error particularly since the quality measures being used are not highly correlated.[46] There may be somewhat more systematic bias by state since even small differences in the directionality of the error within a facility or across facilities in a state can compromise the validity of quality measures substantially.[23] However, there are various problems with the current quality measures such as the stability of the measures, the lack of correlation amongst them and the limited level of risk adjustment that have a far greater effect on the meaningfulness and performance of the quality measures than the level of error in the data.[47, 48]

Using the MDS for research, policy evaluation and planning has the advantage of not requiring the same level of precision as is needed to justify a clinical decision about an individual resident nor even as definitive as should be necessary to publish the relative ranking of one home over another on a given quality measure. Furthermore, there are statistical means of “adjusting” out the idiosyncratic measurement error that can occur in some facilities and not others, still making it possible to examine the effect of states’ policies on resident adjusted outcomes such as pain or ADL.[49] Evidence of the strong monotonic relationship between the CHESS scale and one year mortality among new NH admissions is clearly at least as strong as the Charlson Index or the Deyo-Elixhauser scale as applied to hospital discharge diagnoses. The existence of standardized physical, cognitive, emotional, social and behavioral functioning scales which are largely invariant across
Minimum Data Set Consistency and Validity

different types of patients provides further evidence of the utility of these data for research and policy applications.

**Implications for the MDS 3.0**

The transition from the MDS 2.0 to the 3.0 is a major change since many of the individual items have been altered in important ways and most importantly under the new version the residents’ perspective is supposed to be “heard” if at all possible. Because of the requirement to interview the resident, the MDS 3.0 introduces a new and important feature into the data but one which carries with it new and different kinds of risks. It will be critical to document and systematically characterize residents who could and could not respond to questions and to monitor how that varies across facilities and over time since this aspect of the MDS 3.0 represents an important new kind of measurement that must be considered in comparing the related quality measures. The importance of separating these two kinds of measurements, self-report and staff rating is rooted in the literature that has repeatedly noted that combining self-report and caregiver responses (regardless who the caregiver is – although there may be different issues as a function of the type of caregiver in question) is very dangerous since patients’ and caregivers’ perspective on most aspects of symptom and psycho-social experience are only moderately correlated at best.[50] Thus, for constructs like pain and mood state, which are acknowledged to have basic validity and measurement issues in the current version of the MDS, the MDS 3.0 alters the wording but also introduces additional complexity by capturing separate voices, the patients’ and the staff assessors’, regarding the construct being measured. This additional heterogeneity introduces an expansive new research agenda for those interested in long term care and in basic measurement issues that are the building blocks of an increasing number of quality measurement initiatives.

The introduction of a new MDS, particularly since it requires a mandatory discharge assessment, creates the opportunity to significantly improve the data quality of the MDS discharge records received. In spite of the increasing interest in the use of re-hospitalization, or avoidable re-hospitalizations, as a quality measure, our
Minimum Data Set Consistency and Validity

finding that facilities submit many more discharge records indicating re-hospitalizations than there are Medicare hospital claims, suggests that the MDS discharge record as currently completed and filed, precludes its use as the basis for monitoring this outcome; if a Medicare hospital claim is present, an MDS discharge is likely to be present, but the opposite is not the case. Requiring that a discharge assessment of the patient be completed rather than merely documentation of the discharge will potentially reduce the number of unnecessary discharge record submissions. All dynamic record systems require ongoing and careful monitoring or those responsible for completing the forms will gradually diverge and make their own internally consistent rules that might diverge substantially from those operating in other areas. Any greater specificity regarding when a discharge is recorded and submitted will be a great improvement over what exists today. Future research examining the completeness of these records relative to Medicare claims and enrollment files will hopefully reveal greater correspondence and completeness with relatively little inter-state or facility variation.

In summary, after years of use, our analyses covering virtually all MDS data completed in the US between 1999 and 2007 find that the MDS data cross-walk reasonably well with Medicare hospital claims diagnosis data, both with respect to the match rate and the validity of the MDS diagnoses. The MDS data do seem to be internally consistent and several of the multi-item scales based upon MDS items have moderate levels of internal consistency and reliability, although not quite at the level of research quality data. On balance, the MDS data are very useful for research and program planning and evaluation and the introduction of the MDS 3.0 offers considerable opportunities to improve the quality and completeness of some of the data while at the same time it creates additional analytic challenges precisely because it endeavors to introduce the patient’s voice into this assessment system. Over the next several years as analysts work to establish data quality benchmarks as well as the quality measure distributions using the MDS 3.0, the data presented here on almost a decade of MDS 2.0 data provides a standard of consistency, reliability and completeness against which the MDS 3.0 should be compared.
Minimum Data Set Consistency and Validity
References


Minimum Data Set Consistency and Validity


Figure 1

Facility Variation in the Positive Predictive Value of Medicare Hospital Claims for Heart Failure and MDS admission assessments indicating a diagnosis of Congestive Heart Failure
Figure 2

One Year Survival by CHESS Score Level and Age
Figure 3

<table>
<thead>
<tr>
<th>Year</th>
<th>1999</th>
<th>2000</th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min</td>
<td>65.3</td>
<td>70.6</td>
<td>76.5</td>
<td>79.3</td>
<td>80.9</td>
<td>82.0</td>
<td>80.7</td>
<td>81.4</td>
<td>80.2</td>
</tr>
<tr>
<td>Mean</td>
<td>81.0</td>
<td>83.3</td>
<td>87.2</td>
<td>88.5</td>
<td>89.1</td>
<td>89.5</td>
<td>89.9</td>
<td>89.9</td>
<td>88.8</td>
</tr>
<tr>
<td>Max</td>
<td>89.8</td>
<td>90.7</td>
<td>93.6</td>
<td>94.9</td>
<td>94.7</td>
<td>94.5</td>
<td>94.5</td>
<td>94.4</td>
<td>93.3</td>
</tr>
</tbody>
</table>
Minimum Data Set Consistency and Validity

Table 1
Match Rate between MDS records and Medicare’s Enrollment File from the CMS MDS Registry by Beneficiary age and calendar year

<table>
<thead>
<tr>
<th>Year</th>
<th># with an MDS under 65</th>
<th># with an MDS 65 and over</th>
<th># with an MDS but no age</th>
<th># Match MDS-Mcare under 65</th>
<th># Match MDS-Mcare 65 and over</th>
<th>Percent match under 65</th>
<th>Percent match 65 and over</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999</td>
<td>392,996</td>
<td>2,895,632</td>
<td>1,137</td>
<td>200,907</td>
<td>2,795,080</td>
<td>51.12</td>
<td>96.53</td>
</tr>
<tr>
<td>2000</td>
<td>401,698</td>
<td>2,873,778</td>
<td>861</td>
<td>205,288</td>
<td>2,772,442</td>
<td>51.11</td>
<td>96.47</td>
</tr>
<tr>
<td>2001</td>
<td>419,180</td>
<td>2,907,589</td>
<td>8</td>
<td>213,819</td>
<td>2,787,048</td>
<td>51.01</td>
<td>95.85</td>
</tr>
<tr>
<td>2002</td>
<td>435,066</td>
<td>2,892,291</td>
<td>21</td>
<td>221,537</td>
<td>2,782,532</td>
<td>50.92</td>
<td>96.21</td>
</tr>
<tr>
<td>2003</td>
<td>456,343</td>
<td>2,876,013</td>
<td>7</td>
<td>232,246</td>
<td>2,778,992</td>
<td>50.89</td>
<td>96.63</td>
</tr>
<tr>
<td>2004</td>
<td>474,974</td>
<td>2,880,140</td>
<td>21</td>
<td>242,819</td>
<td>2,787,829</td>
<td>51.12</td>
<td>96.79</td>
</tr>
<tr>
<td>2005</td>
<td>495,605</td>
<td>2,926,867</td>
<td>47</td>
<td>256,551</td>
<td>2,844,360</td>
<td>51.77</td>
<td>97.18</td>
</tr>
<tr>
<td>2006</td>
<td>506,175</td>
<td>2,932,121</td>
<td>41</td>
<td>265,438</td>
<td>2,841,366</td>
<td>52.44</td>
<td>96.90</td>
</tr>
<tr>
<td>2007</td>
<td>520,368</td>
<td>2,945,183</td>
<td>82</td>
<td>274,922</td>
<td>2,858,947</td>
<td>52.83</td>
<td>97.07</td>
</tr>
<tr>
<td>2008</td>
<td>536,114</td>
<td>2,991,047</td>
<td>19</td>
<td>283,471</td>
<td>2,855,297</td>
<td>52.88</td>
<td>95.46</td>
</tr>
</tbody>
</table>
### Table 2

#### Table 2. General Characteristics of Population by Year

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n=790,227</td>
<td>n=790,617</td>
<td>n=773,746</td>
<td>n=718,555</td>
</tr>
<tr>
<td>Mean age (±SD)</td>
<td>81.1±7.3</td>
<td>81.1±7.3</td>
<td>81.0±7.3</td>
<td>81.0±7.4</td>
</tr>
<tr>
<td>65-74</td>
<td>20.3%</td>
<td>20.5%</td>
<td>20.7%</td>
<td>21.1%</td>
</tr>
<tr>
<td>75-84</td>
<td>45.7%</td>
<td>46.1%</td>
<td>46.1%</td>
<td>45.6%</td>
</tr>
<tr>
<td>85+</td>
<td>33.9%</td>
<td>33.4%</td>
<td>33.1%</td>
<td>33.3%</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>33.4%</td>
<td>34.5%</td>
<td>35.1%</td>
<td>35.7%</td>
</tr>
<tr>
<td>Female</td>
<td>66.6%</td>
<td>65.5%</td>
<td>64.9%</td>
<td>64.3%</td>
</tr>
<tr>
<td>Mean number of Dx (±SD)</td>
<td>7.0±2.2</td>
<td>7.4±2.1</td>
<td>7.7±1.9</td>
<td>8.0±1.8</td>
</tr>
<tr>
<td>1-2</td>
<td>3.8%</td>
<td>2.7%</td>
<td>1.9%</td>
<td>1.3%</td>
</tr>
<tr>
<td>3-4</td>
<td>12.5%</td>
<td>9.7%</td>
<td>7.4%</td>
<td>5.8%</td>
</tr>
<tr>
<td>5-6</td>
<td>20.5%</td>
<td>17.3%</td>
<td>14.4%</td>
<td>11.8%</td>
</tr>
<tr>
<td>7-8</td>
<td>21.6%</td>
<td>20.8%</td>
<td>18.8%</td>
<td>16.4%</td>
</tr>
<tr>
<td>9</td>
<td>41.6%</td>
<td>49.6%</td>
<td>57.5%</td>
<td>64.7%</td>
</tr>
</tbody>
</table>

SD: standard deviation
Table 3

Positive Predictive Value of MDS based diagnosis relative to ICD-9 Diagnoses on the Medicare Hospital Claim for selected Diagnoses

<table>
<thead>
<tr>
<th>Medicare Hospital Claim Diagnoses</th>
<th>MDS Diagnoses</th>
<th>2006</th>
<th>Inter-quartile range 2006</th>
<th>25th percentile</th>
<th>75th percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any hypertension</td>
<td>Hypertension</td>
<td>0.61</td>
<td>0.038</td>
<td>0.591</td>
<td>0.628</td>
</tr>
<tr>
<td>Heart failure/cardiomegaly</td>
<td>Heart failure</td>
<td>0.78</td>
<td>0.038</td>
<td>0.760</td>
<td>0.798</td>
</tr>
<tr>
<td>Cerebrovascular disorders</td>
<td>Stroke/TIA</td>
<td>0.32</td>
<td>0.064</td>
<td>0.299</td>
<td>0.363</td>
</tr>
<tr>
<td>Parkinson's disease</td>
<td>Parkinson's disease</td>
<td>0.60</td>
<td>0.090</td>
<td>0.573</td>
<td>0.663</td>
</tr>
<tr>
<td>Alzheimer's disease</td>
<td>Alzheimer's disease</td>
<td>0.66</td>
<td>0.076</td>
<td>0.629</td>
<td>0.705</td>
</tr>
<tr>
<td>Brain degeneration</td>
<td>Any-type dementia</td>
<td>0.32</td>
<td>0.058</td>
<td>0.286</td>
<td>0.344</td>
</tr>
<tr>
<td>Asthma/COPD/emphysema</td>
<td>Asthma/COPD/emphysema</td>
<td>0.80</td>
<td>0.036</td>
<td>0.778</td>
<td>0.815</td>
</tr>
<tr>
<td>Any pneumonia</td>
<td>Pneumonia</td>
<td>0.63</td>
<td>0.064</td>
<td>0.613</td>
<td>0.677</td>
</tr>
<tr>
<td>Depressive disorders</td>
<td>Depression</td>
<td>0.25</td>
<td>0.050</td>
<td>0.232</td>
<td>0.283</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>Diabetes mellitus</td>
<td>0.69</td>
<td>0.039</td>
<td>0.679</td>
<td>0.718</td>
</tr>
<tr>
<td>Any cancer</td>
<td>Cancer</td>
<td>0.55</td>
<td>0.076</td>
<td>0.529</td>
<td>0.606</td>
</tr>
<tr>
<td>Any anemia</td>
<td>Anemia</td>
<td>0.51</td>
<td>0.055</td>
<td>0.495</td>
<td>0.551</td>
</tr>
<tr>
<td>Any UT infection</td>
<td>Urinary tract infection</td>
<td>0.62</td>
<td>0.054</td>
<td>0.591</td>
<td>0.645</td>
</tr>
</tbody>
</table>
Table 4
Positive Predictive Value and Internal Consistency of selected MDS items on the Admission MDS for 2006 and inter-quartile range across states

<table>
<thead>
<tr>
<th>MDS &quot;Gold&quot;</th>
<th>MDS Diagnoses</th>
<th>2006</th>
<th>Inter-quartile Range 2006</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADL(a) ≠ 0</td>
<td>Hemiplegia</td>
<td>0.99</td>
<td>0.008</td>
</tr>
<tr>
<td>ADL(a) ≠ 0</td>
<td>Amputation</td>
<td>0.99</td>
<td>0.010</td>
</tr>
<tr>
<td>ADL(a) ≠ 0</td>
<td>Bed-ridden</td>
<td>0.99</td>
<td>0.013</td>
</tr>
<tr>
<td>ADL(a) ≠ 0</td>
<td>Bed mobility = 0</td>
<td>0.97</td>
<td>0.000</td>
</tr>
<tr>
<td>ADL(a) ≠ 0</td>
<td>Pressure sore stage 3-4</td>
<td>0.99</td>
<td>0.015</td>
</tr>
<tr>
<td>CPS</td>
<td>Alzheimer's disease</td>
<td>0.96</td>
<td>0.022</td>
</tr>
<tr>
<td>CPS</td>
<td>Vascular-type dementia</td>
<td>0.93</td>
<td>0.029</td>
</tr>
<tr>
<td>Communication=Rarely/Never understood</td>
<td>Aphasia</td>
<td>0.23</td>
<td>0.072</td>
</tr>
<tr>
<td>Communication=Usually understood</td>
<td>Aphasia</td>
<td>0.26</td>
<td>0.063</td>
</tr>
<tr>
<td>Vision impairment</td>
<td>Cataract</td>
<td>0.47</td>
<td>0.115</td>
</tr>
<tr>
<td>Edema</td>
<td>No dehydration</td>
<td>0.38</td>
<td>0.139</td>
</tr>
<tr>
<td>Joint pain</td>
<td>Arthritis</td>
<td>0.42</td>
<td>0.071</td>
</tr>
<tr>
<td>Cancer</td>
<td>Chemotherapy</td>
<td>0.57</td>
<td>0.137</td>
</tr>
<tr>
<td>Asthma/COPD/emphysema</td>
<td>Respirator/oxygen</td>
<td>0.37</td>
<td>0.013</td>
</tr>
<tr>
<td>Any ulcer</td>
<td>Ulcer care</td>
<td>0.98</td>
<td>0.008</td>
</tr>
<tr>
<td>Edema</td>
<td>Diuretic received</td>
<td>0.55</td>
<td>0.132</td>
</tr>
</tbody>
</table>
Table 5  

Standardized Alpha Multi-Item Scale Reliability & Internal Consistency  
Stratified by Median Activities of Daily Living²

<table>
<thead>
<tr>
<th></th>
<th>ADL-high</th>
<th>ADL-low</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Social engagement scale³</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPS-high</td>
<td>0.62</td>
<td>0.60</td>
</tr>
<tr>
<td>CPS-low</td>
<td>0.60</td>
<td>0.59</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>ADL-high</th>
<th>ADL-low</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mood scale⁴</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPS-high</td>
<td>0.66</td>
<td>0.65</td>
</tr>
<tr>
<td>CPS-low</td>
<td>0.59</td>
<td>0.57</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>ADL-high</th>
<th>ADL-low</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Behavior scale⁵</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPS-high</td>
<td>0.53</td>
<td>0.53</td>
</tr>
<tr>
<td>CPS-low</td>
<td>0.38</td>
<td>0.39</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>ADL-high</th>
<th>ADL-low</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pain⁶</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPS-high</td>
<td>0.49</td>
<td>0.52</td>
</tr>
<tr>
<td>CPS-low</td>
<td>0.47</td>
<td>0.49</td>
</tr>
</tbody>
</table>

² The numbers in the cell indicate the standardized Cronbach Coefficient Alpha.  
³ Includes a sum of 6 variables: At ease interacting with others; At ease with structured activities; At ease self-initiated activities; Established own goals; Pursues involvement in facility; Accepts invites to most activities.  
⁴ Includes a sum of 10 variables: Verbal expression of distress; Repetitive Verbalization; Self Deprecation; Unrealistic Fears; Recurrent thoughts of death; Repetitive Health Complaints; Crying Tearfulness; Repetitive Physical Movements; Leaves 25% food uneaten; Mood persistence.  
⁵ Includes a sum of 4 variables: Verbally abusive; Physically abusive; Socially inappropriate; Manic depression (bipolar disease).  
⁶ Includes a sum of 2 variables: Frequency of pain; Intensity of pain