



Published in final edited form as:

J Am Med Dir Assoc. 2011 January ; 12(1): 38–43. doi:10.1016/j.jamda.2010.02.001.

Validation of the Minimum Data Set in identifying hospitalization events and payment source

Shubing Cai, PhD¹, Dana B. Mukamel, PhD², Peter Veazie, PhD³, and Helena Temkin-Greener, PhD³

Shubing Cai: Shubing_Cai@brown.edu; Dana B. Mukamel: dmukamel@uci.edu; Peter Veazie: Peter_Veazie@urmc.rochester.edu; Helena Temkin-Greener: Helena_Temkin-Greener@urmc.rochester.edu

¹ Center for Gerontology and Health Care Research, The Warren Alpert Medical School, Brown University, 121 S. Main Street, Box G-S121 (6), Providence, RI 02912, Phone: 401-863-9586; fax: 401-863-3489

² Department of Medicine, Center for Health Policy Research, University of California, Irvine, 111 Academy, Suite 220, Irvine CA 92697-5800, Phone: 949-824-8873; fax: 949-824-3388

³ Department of Community and Preventive Medicine, University of Rochester School of Medicine, Box 644, 601 Elmwood Avenue, Rochester, New York 14642, Phone: 585-273-5464 (Peter Veazie), 585-275-8713 (Helena Temkin-Greener); fax: 585-461-4532

Keywords

MDS; nursing home; data quality; hospitalization; payment source

INTRODUCTION

Quality of care in US nursing homes (NH) remains an on-going concern and a focus of much current research. The Minimum Data Set (MDS) is a major source of information used in this research. Required by federal regulation, the MDS assessment has to be completed at the admission and quarterly thereafter if a resident is still in the NH. Furthermore, NHs are also required to submit additional assessments for residents whose care is covered by Medicare¹. The MDS contains much information about the residents demographics, health conditions, preferences, treatments, sources of payment, and about transfers to other care settings, such as hospitals. Many data elements contained in the MDS are generally considered to be reliable^{2–4}. However, to the best of our knowledge, information on hospitalization events and on payment sources has not been validated.

Researchers who have focused on hospitalizations of nursing home residents usually link Medicare claims data with the MDS to identify hospitalization events^{5–6}. Researchers who are interested in the impact of payer status on quality of nursing home care need either to restrict their studies to only seven states, which are assumed to have reliable payer source information in the MDS⁷, or use alternative data sources (i.e. MEPS data)^{8, 9} that are not as comprehensive as the MDS.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

While linkages of multiple data sources are feasible, they are costly and time consuming. In addition, linking multiple datasets is likely to cause reductions in the sample size, because records cannot always be matched across data sets. Furthermore, the loss of records during such merges may introduce a bias if the records that do not link are different in any systematic way from those that are linked. Linkage with certain sources, for example with Medicare data, to identify hospitalization events will also restrict the study sample to only include residents enrolled in Medicare fee-for-service (FFS). Furthermore, time lag in the availability of the data often prevents timely linkage. For example, while the time lag for the MDS data are relatively short, the lag for the receipt of claims data is much longer, especially for Medicaid claims, which can take years for the data to be available¹⁰. If the MDS data alone were sufficiently reliable to identify hospitalization events or payment source, such data linkages would not be necessary. The purpose of this study is to examine the quality of hospitalization records and payment source within the MDS data.

METHODS

Data

This analysis is based on data for California (CA), New York (NY), Ohio (OH), and Texas (TX) in calendar year 2003. These four states are chosen because they represent the largest nursing home populations and are diverse with regard to nursing home practice and policies. The following datasets are used in this analysis: the Minimum Data Set (MDS), Medicare Provider Analysis and Review File (MedPAR), Medicare denominator file, Medicaid Analytic Extract (MAX) long-term care (LTC) file and MAX personal summary file.

As already discussed, the MDS is a federal mandated nursing home assessment tool containing detailed information about health status for all residents in Medicare/Medicaid certified nursing facilities.

Medicare Denominator File includes demographic and enrollment information for Medicare beneficiaries in each calendar year and is used here to identify the eligible study population. MedPAR data are extracted from Medicare claims, and each record represents an inpatient or skilled nursing facility (SNF) stay¹¹.

Medicaid Analytic Extract personal summary file and long-term care file are based on state Medicaid claims extracted by CMS. Personal Summary File contains information, such as demographic characteristics, enrollment status, and utilization summaries, for every individual who enrolled in Medicaid during the year. LTC File is comprised of claims for institutional long term care services provided by nursing facilities and intermediate care facilities. The LTC file includes information about individual's LTC stay, such as facility type, starting and ending dates of service, and discharge status¹².

Analytic Approach

Identification of sample population—We first linked the MDS data with the Medicare denominator file to identify nursing home residents who were eligible for Medicare Part A from the beginning of the year. Beneficiaries who were enrolled in Medicare managed care were excluded because Medicare claims for managed care enrollees may not be complete. To validate payment status, the linked datasets, as described above, were further merged with the MAX personal summary and LTC files. Medicaid managed care enrollees were also excluded due to the possibility of incomplete claims.

Data records were linked by social security number (SSN), health insurance code (HIC), and date of birth. We only kept individual records that had unique matches (e.g. if one HIC matched two different SSN, the record was excluded).

In the following section, we discuss the identification of hospitalization events and payment sources. We first identified this information based on the MDS data. We then linked the MDS and claims data to determine the same information from the external claims data, and then compare these two methods. Since Medicare and Medicaid claims data are originally used for reimbursement purposes, identification using these data should be more accurate than by the MDS alone. Therefore, we consider identification with claims data to be the “gold standard”.

Identification of hospitalization events—Hospitalization events identified via the MDS are based on discharge destination codes for ‘acute hospital’, ‘psychiatric hospital’, or ‘rehabilitation hospital.’ Correspondingly, we selected hospitalization admissions in MedPAR with service types identified as “acute hospital”, “rehabilitative hospital” or “psychiatric hospitals”. These selected admissions were then linked with the MDS data to identify whether they originated from a nursing facility. A MedPAR hospital admission was not considered as originating from a nursing home if any of the following conditions were met: (1) MedPAR hospital admission was either prior to the nursing home admission or 3 days after an MDS discharge date. We used a 3-day criterion to allow for some degree of recording inaccuracy. (2) There was no discharge record in the MDS and the interval between hospital admission in MedPAR and most recent MDS assessment exceeded 100 days (CMS requires at least one assessment per quarter; we used 100 days as a conservative measure). (3) When there were consecutive hospital admissions in MedPAR but there were no MDS records, or MedPAR SNF records interspersed between them (indicating no nursing home stays between consecutive admissions), only the first hospital admission was considered as originating from the nursing home.

MDS discharges were considered to be matched with MedPAR hospital admissions if both of the following were met: (1) MedPAR hospital admission was within 3 days of MDS discharge, and (2) MDS discharge type matched the MedPAR admission type. We only validated hospitalizations beginning in the second quarter of CY2003, or later in that year to correctly identify nursing home-originating hospitalizations. If a hospital admission occurred before the first available MDS assessment in the first quarter, it would be hard to determine whether it originated from a nursing home (e.g. there may have been an MDS assessment in CY2002) or from another setting. Starting with the second quarter allows us to assess the setting from which a MedPAR hospital admission may have originated. The identification processes is summarized in the Figure 1.

Identification of payment source—The payment source for all MDS assessments was evaluated by two methods: solely by the MDS and combination of the MDS and claims data. Our analysis was based on assessments rather than on individuals since payment status for an individual could change over the course of a nursing home stay. The payment source for each MDS assessment was first determined solely by two MDS variables (“reasons for assessment”-AA8b, and “current payment source”-AA7a). The payment source was determined consecutively, starting with Medicare. The assessment was considered to be under Medicare if the primary reason for the assessment was Medicare prospective payment (according to AA8b). For all non-Medicare assessments, the payment status was then determined by the ‘current payment source’ variable (AA7a) in the MDS. However in the MDS, item ‘current payment source’ was not evaluated at every assessment, which resulted in missing values. Therefore we replaced the missing payment source with payment source of the previous assessment (if it was not missing in the previous assessment), assuming nursing home might not update the payment source if there was no change in payment status. This was the only MDS variable based from which Medicaid or private payment status may be determined. The payment source was considered to be Medicaid if item AA7a

was ‘Medicaid per diem’. We grouped all non-Medicare & non-Medicaid payment sources as ‘other/private.’

We then identified payment source for each MDS assessment by linking the MDS and claims data. Medicare payment status was jointly determined by MedPAR SNF claims and the MDS data. Multiple items in MedPAR SNF claims (i.e. beginning and ending date of the service, ending date of Medicare coverage for SNF stay, Medicare exhaust date, and the number of covered days of care chargeable to Medicare) were used to identify the Medicare reimbursement duration. These identified Medicare reimbursement periods were then linked with the MDS assessments for each individual. MDS assessment was considered Medicareable if the assessment date was within the MedPAR SNF period. For all non-Medicare assessments, if the assessment date was after the beginning date of Medicaid service (identified by MAX claims), the assessment was considered under Medicaid payment status. Otherwise, payment status was deemed as ‘other/private’. The identification process is summarized in Figure 2.

Comparing two types of identifications: MDS versus claims data—Using the “gold standard” identification defined above, we calculated false negative ($100 \times [1 - \text{sensitivity}]$) and false positive error rates ($100 \times [1 - \text{positive predictive value}]$) to assess the accuracy of the MDS alone in identifying hospital admissions and payment source. Sensitivity was measured as the percentage of payment source and nursing home-originating hospitalizations, derived from claims data, which could be matched with the payment source and hospitalization events identified by the MDS. Positive predictive value (PPV) was measured as the percentage of MDS-identified payment source/hospitalization events that could be validated by claims data.

RESULTS

Validation of hospitalization events

Our study sample contained 207,554 MDS hospital-related discharges and 204,975 MEDPAR nursing home-originating hospital admissions during the observational period. Overall, 86% of the MDS discharges were matched with MedPAR admissions using the criteria described above. As shown in Table 1, the false negative error rate, indicating the sensitivity of the MDS data, varied across the four states, ranging from 6.8% in Ohio to 19.5% in Texas. The false positive error rates, indicating the likelihood that MDS falsely identifies hospitalizations, i.e. those not reported in MedPAR, were quite similar across the four states, ranging from 12.0% in NY to 15.7% in TX.

Furthermore, we examined the MDS hospital-related discharges that could not be matched with MedPAR. As indicated in Table 2, 5.9% of the unmatched MDS records in NY were due to the lack of agreement between the MDS hospital-related discharge and the MEDPAR admission service type (i.e. psychiatric hospitals related transfer in the MDS, but long-term care hospital in MedPAR), even though hospital-related discharge dates in MDS were matched to MEDPAR admission dates. This was also the reason for the 19.0% unmatched records in TX. If we were to count this type of unmatched MDS records as matches, then the false positive error rates of the MDS would decrease, i.e. from 15.7% to 12.7% in TX (Table 1 column 5).

Some of these misspecifications might be ascribed to emergency room (ER) transfers or deaths. For example, a number of residents who had hospital-related MDS discharge but no corresponding MEDPAR admissions appeared to have returned to the nursing home within two days of discharge. As presented in Table 2, this explains 12.3% to 18.8% (in NY and TX, respectively) of unmatched MDS hospitalization-related discharges across the four

states. In all likelihood these individuals were discharged to ER but were not subsequently admitted for inpatient care. MedPAR does not contain claims for ER if an individual is not eventually hospitalized¹¹. Furthermore, between 11.0% and 14.9% of the unmatched records could be attributed to residents who died, perhaps also in the ER, within two days of being discharged to a hospital (Table 2). These circumstances, however, could not fully explain the unmatched identification between the MDS and MedPAR data.

Validation of payment source

In total 1,801,717 MDS assessments were obtained across the four states for the purpose of payment source validation. The overall agreement rate between the two methods of identification (MDS alone versus combined MDS and claims data) was 77%. As shown in Table 3, the overall agreement rate was highest in OH (93%), and lowest in TX (69%). In identifying Medicare payment source, the MDS exhibited low false negative rates: ranging from 0.4% in OH to 1% in NY. This suggests that virtually all assessments with the Medicare SNF payer status could be identified with the MDS data alone. However, the false positive rate of identifying Medicare status with the MDS data was relatively high: ranging from 6% in OH to 15% in TX. Thus it appears that 6%–15% of the MDS Medicare assessments (MDS item AA8b) did not have corresponding Medicare reimbursement claims in these four states.

With regard to identifying Medicaid payment source, the false positive rate of relying on the MDS alone was low: 2% in CA, NY, and OH, and 3% in TX. However, the false negative rate in using the MDS data was much higher: 50% in NY, 52% in CA, 55% in TX, and 11% in OH. Although MDS recording of payment status as Medicaid is quite accurate, a large proportion of MDS assessments that also should be recorded as Medicaid were misspecified into other payment categories. As presented in Table 4, this misspecification was due to a high percentage of missing records for current payment source and a misclassification of Medicaid payment source into private payment or Medicare.

DISCUSSION

Unnecessary hospitalizations and potentially unequal care among residents with different payer sources are two important issues in nursing home care. Unnecessary and unequal care may result in negative health outcomes for the residents and lead to excessive financial costs to society^{13–18}. How to reduce unnecessary care and its subsequent costs, without diminishing the quality of care, i.e. how to provide the most efficient care, is especially of concern in the current political environment and the ongoing efforts to reform health care system.

In order to investigate the causes and solutions for these issues it is imperative to understand their true magnitude; and to accomplish this it is important to correctly identify the key variables (hospitalization events and payer source). While linking the MDS with Medicare and Medicaid claims data provides a more accurate way to identify these variables, it is costly, time-consuming, and has its own limitations as previously discussed. Therefore, this study was undertaken to evaluate the accuracy of the MDS data in identifying these two variables, and to provide researchers with the necessary information to decide which data sources to use, given the research question.

Unfortunately, the MDS does not seem to be an ideal source for identifying either payer source or hospitalization events without supplemental information from claims data. We find modest levels of misspecification in the MDS with regard to the identification of hospitalization events. Although some of misspecifications could be explained, but the main reasons for the misspecification are left unknown. As to identifying payment source, the

MDS has a low false negative rate but relatively high false positive rate. This is quite expected since according to the federal regulations, nursing homes have to submit sequential Medicare assessments (at 5, 14, 30 and 60 days) in order to get paid under the Medicare prospective payment system. This regulation provides an important financial incentive for nursing homes to accurately submit all assessments required for payment. However, even though nursing homes submit MDS assessments for Medicare prospective reimbursement, not all of these MDS Medicare-related assessments are actually reimbursed by Medicare, per claims data. This leads to false positive identifications of Medicare payment status. The false positive rate in TX is higher than that in the other three states. The MDS especially have poor quality in identifying Medicaid and private payment sources, as demonstrated by the high false negative rates. Therefore linkage of the MDS and MAX data seems to be the only way to accurately ascertain these payment sources.

However, although the MDS alone is not ideal for identifying hospitalization and payment source for nursing home residents, it may still be sufficient in some cases. For example, if the research is mainly focused on the occurrence of transferring residents rather than the reasons for transfer, the MDS alone may still be informative. Moreover, relying on the MDS alone may be reasonable if the research interest is to investigate the within-facility difference in hospitalization rates among residents by race/ethnicity, since there is no reason to believe the misspecification will be systematically different among residents within a facility. In addition, the MDS seems to be the only source for investigating hospitalization events among Medicare population who enrolled in Managed care since Medicare claims data do not contain hospitalization records for this population. As for identifying payer source, although the MDS could not differentiate the Medicaid from the private-pay source, the MDS data seems to be reasonable in determining Medicare payment status. Therefore, if the research interest is only focused on nursing home residents who receive post-acute/rehabilitation care, the MDS may be a sufficient resource, at least in selected states. In addition, the payer source information in Ohio seems to be better than the other states, thus it may be a reasonable source for some research purposes.

Some limitations of this study should be mentioned. Our evaluation is based on data from four states in CY2003. It may not be prudent to generalize from these results to other states and years. Data quality varies from state to state and may also have changed substantially following the 2003 implementation of the CMS nursing home compare website. Similar analyses for other states and across time may shed additional information.

CONCLUSION

Our study provides new information about the accuracy of the MDS data. Moreover, this is the first study to examine the accuracy of the MDS in identifying payment source by linking the MDS and Medicaid claims data together. Overall, the MDS does not seem to be an **ideal** source for identifying either hospitalization events or payment source without linking with claims data, although it may still be useful in some cases. The decision on whether to link MDS with MedPAR or use MDS alone to identify nursing home-originating hospitalizations should depend on the specific research questions at hand.

Acknowledgments

We gratefully acknowledge financial supporting from the National Institute on Aging, Grant R01 AG23077.

References

1. CMS. The MDS 2.0. [Accessed on Dec 28, 2009].
http://www.cms.hhs.gov/NursingHomeQualityInits/01_overview.asp

2. Hawes C, Morris JN, Phillips CD, et al. Development of the nursing home Resident Assessment Instrument in the USA. *Age Ageing* Sep;1997 26 (Suppl 2):19–25. [PubMed: 9464550]
3. Mor V, Angelelli J, Jones R, et al. Inter-rater reliability of nursing home quality indicators in the U.S. *BMC. Health Serv Res* Nov 4;2003 3(1):20.
4. Morris, J.; Moore, T.; Jones, R., et al. *Validation of Long-term and Post-acute Quality Indicators*. Baltimore, MD: Centers for Medicare and Medicaid services; 2003.
5. Carter MW. Variations in hospitalization rates among nursing home residents: the role of discretionary hospitalizations. *Health Serv Res* Aug;2003 38(4):1177–1206. [PubMed: 12968823]
6. Intrator O, Schleinitz M, Grabowski DC, et al. Maintaining continuity of care for nursing home residents: effect of states' Medicaid bed-hold policies and reimbursement rates. *Health Serv Res* 2009;44(1):33–55. [PubMed: 18783452]
7. Grabowski DC, Gruber J, Angelelli J. Nursing home quality as a common good. *The Review of Economics and Statistics* 2008;90(4):754–764. [PubMed: 20463859]
8. Konetzka RT, Spector W, Shaffer T. Effects of nursing home ownership type and resident payer source on hospitalization for suspected pneumonia. *Med Care* Oct;2004 42(10):1001–1008. [PubMed: 15377933]
9. Troyer JL. Examining differences in death rates for medicaid and non-medicaid nursing home residents. *Med Care* Oct;2004 42(10):985–991. [PubMed: 15377931]
10. Research Data Assistance Center(ResDac). [Accessed on Dec 28, 2009]. Available at: http://www.resdac.umn.edu/Available_CMS_Data.asp
11. Research Data Assistance Center(ResDac). Medicare Data File Description. [Accessed on Dec 28, 2008]. Available at: http://www.resdac.umn.edu/Medicare/data_file_descriptions.asp#rif
12. Research Data Assistance Center(ResDac). the Medicaid Analytic Extract (MAX) File Description. [Accessed on Dec 28, 2008]. Available at: http://www.resdac.umn.edu/Medicaid/data_documentation.asp
13. Mor V, Zinn J, Angelelli J, et al. Driven to tiers: socioeconomic and racial disparities in the quality of nursing home care. *Milbank Q* 2004;82(2):227–256. [PubMed: 15225329]
14. Grabowski DC, O'Malley AJ, Barhydt NR. The costs and potential savings associated with nursing home hospitalizations. *Health Aff (Millwood)* Nov–Dec;2007 26(6):1753–1761. [PubMed: 17978395]
15. Kruse RL, Mehr DR, Boles KE, et al. Does hospitalization impact survival after lower respiratory infection in nursing home residents? *Med Care* Sep;2004 42(9):860–870. [PubMed: 15319611]
16. Loeb M, Carusone SC, Goeree R, et al. Effect of a clinical pathway to reduce hospitalizations in nursing home residents with pneumonia: a randomized controlled trial. *Jama* Jun 7;2006 295(21):2503–2510. [PubMed: 16757722]
17. Ouslander JG, Weinberg AD, Phillips V. Inappropriate hospitalization of nursing facility residents: a symptom of a sick system of care for frail older people. *J Am Geriatr Soc* Feb;2000 48(2):230–231. [PubMed: 10682957]
18. Binder EF, Kruse RL, Sherman AK, et al. Predictors of short-term functional decline in survivors of nursing home-acquired lower respiratory tract infection. *J Gerontol A Biol Sci Med Sci* Jan; 2003 58(1):60–67. [PubMed: 12560413]

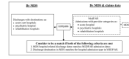


Figure 1.
Identification of hospitalization events

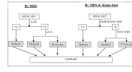


Figure 2.
Identification of payment source

Table 1

Agreement between MDS and MEDPAR on hospitalizations of nursing home residents

State	MDS Hospital Related Discharges	False negative error rate(%) ^a	False positive error rate(%) ^b	False positive error rate(%) ^c
California	48,785	11.9	13.0	11.8
New York	57,933	10.7	12.0	11.2
Ohio	47,085	6.8	14.2	12.1
Texas	53,751	19.5	15.7	12.7

^a False negative error rate=100(1-sensitivity)

^b False positive error rate=100(1-positive predictive value)

^c MDS discharge date matches MEDPAR admission date, but MEDPAR admissions types are not restricted to the three types in our selection criteria(acute, rehabilitative or psychiatric)

Table 2

Possible Reasons for Lack of MDS to MEDPAR Match

State	MDS discharges unmatched with MEDPAR	Reasons for Lack of MDS to MEDPAR Match ^a				Unexplained unmatched discharges (%) ^a
		MDS discharge destination and MEDPAR admission source do not match (%) ^a	MDS discharge followed by re-entry within 2 days: possible ER visit (%) ^a	Death within 2 days of discharge date (%) ^a		
California	6,336 (13.0)	8.7	15.6	14.9	60.8	
New York	6,929 (12.0)	5.9	12.3	14.5	67.3	
Ohio	6,689 (14.2)	14.8	17.6	11.7	55.9	
Texas	8,469 (15.7)	19.0	18.8	11.0	51.2	

^a as percent of unmatched MDS discharges

Table 3

Accuracy of the MDS in identifying Medicare payment source

	California		New York		Ohio		Texas	
	Medicare	Medicaid	Medicare	Medicaid	Medicare	Medicaid	Medicare	Medicaid
Total number of MDS assessments (% overall accuracy rate⁺)	417,246 (75.84%)		519,799 (70.92%)		423,550 (92.79%)		441,122 (68.58%)	
Number of assessments with payment source identified by MDS	235,365	74,584	251,565	124,276	212,074	146,390	221,056	88,689
Number of assessments with payment source identified by claims data	220,681	151,535	227,180	244,888	199,952	161,617	190,033	192,221
Number of MDS assessments validated by the claims data	219,216	73,114	224,666	122,119	199,130	143,876	188,079	85,874
False negative rate (%)	0.66	51.75	1.11	50.13	0.41	10.98	1.03	55.33
False positive rate (%)	6.86	1.97	10.69	1.74	6.10	1.72	14.92	3.17

⁺Percentage of the MDS assessments that can be validated by claims data

Table 4

The reasons for misspecification of the MDS in identifying Medicaid payer status

Total assessments identified as under Medicaid payer source by claims data	California N=151,535	New York N=244,888	Ohio N=161,617	Texas N=192,221
Due to missing (%)	31.8	31.91	0	30.43
Misclassified into private-pay (%)	15.09	13.04	8.94	14.94
Misclassified into Medicare (%)	4.89	5.19	2.04	9.95