Published in final edited form as:

Med Care. 2008 November ; 46(11): 1184–1187. doi:10.1097/MLR.0b013e31817925d2.

Ability of Medicare Claims Data to Identify Nursing Home Patients:

A Validation Study

Siran M. Koroukian, PhD^{*,†}, Fang Xu, MS^{*}, and Patrick Murray, MD, MS[‡]

^{*}Department of Epidemiology and Biostatistics, School of Medicine, Case Western Reserve University

[†]Cancer Aging Program, Case Comprehensive Cancer Center

[‡]Center for HealthCare Research and Policy, MetroHealth Medical Center, Cleveland, Ohio

Abstract

Background—There are currently no Medicare claims-based algorithms to identify patients receiving nursing home care (NHC). This constitutes an important limitation in outcome studies using population-based data.

Objectives—To assess the ability of claims data to identify patients receiving NHC, using the nursing home Minimum Data Set (MDS) as the gold standard. We hypothesized that physician claims carrying relevant Evaluation and Management (E&M) procedure codes would be an adequate source to identify nursing home patients.

Research Design—Cross-sectional study using the Ohio Cancer-Aging Linked Database, developed by linking records from the Ohio Cancer Incidence Surveillance System with multiple sources of data, including Medicare enrollment and claims files, and the MDS.

Subjects—Patients 65 years of age or older residing in Ohio, and diagnosed with incident breast, prostate, or colorectal cancer during years 1997–2001. We limited our study cohort to fee-for-service patients receiving NHC during calendar year 2002, with a look-back period in the claims data to November 2001.

Measures—Sensitivity and positive predictive value (PPV).

Results—Sensitivity, or the proportion of patients identified through the MDS file who were also successfully identified through the claims data was 88.1%. PPV, or the proportion of patients with the relevant E&M codes who were also identified in the MDS file was 83.9%.

Conclusions—Carrier files may be an acceptable data source to identify nursing home patients, paving the way for future risk adjustment techniques to account for nursing home status.

Keywords

nursing home; Minimum Data Set; Medicare claims data

Copyright © 2008 by Lippincott Williams & Wilkins

Reprints: Siran M. Koroukian, PhD, Department of Epidemiology and Biostatistics, School of Medicine, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106-4945. skoroukian@case.edu..

Cancer incidence data were obtained from the Ohio Cancer Incidence Surveillance System (OCISS), Ohio Department of Health. Use of these data does not imply that the Ohio Department of Health either agrees or disagrees with any presentation, analyses, interpretations, or conclusions. Information about the OCISS may be obtained at odh.state.oh.us/ODHPrograms/CI_SURV/ci_surv1.htm.

Nursing home (NH) residents are the frailest and sickest of elders. With high disease prevalence, as well as functional limitations and/or cognitive impairment,¹ they are highly susceptible to poor health outcomes and death. NH patients represent a small proportion of elders, but they consume a disproportionate share of health care dollars.²

Although risk adjustment techniques using administrative claims data focus on the presence or absence of certain diagnostic codes,^{3,4} they fail to account for the patient's NH status, which reflects a high level of dependence and complexity of care needs. Consequently, comorbidity indices assign the same score to both NH and community-dwelling patients presenting with the same diagnoses in claims data.

For proper risk adjustment, the identification of NH patients is of paramount importance. With regard to cancer-related disparities, for example, findings from a recent study of older breast, colorectal, or prostate cancer patients⁵ showed that patients admitted to a NH before cancer diagnosis were significantly more likely than their community-dwelling counterparts to have their cancer unstaged, or not have enough information on their cancer to indicate a stage,⁶ due to incomplete diagnostic evaluation. Another study⁷ of NH residents diagnosed with breast, colorectal, prostate, or lung cancer confirmed these findings, and further documented high mortality within 3 months of diagnosis, and low rates of surgery, chemotherapy, and/or radiation. To identify NH patients, these studies have resorted to linking state cancer surveillance data with other sources of data, such as the NH Minimum Data Set (MDS) or state Medicaid data.^{8,9}

Because physicians use special Evaluation and Management (E&M) procedure codes when providing care to NH patients, we hypothesized that Carrier claims data would be an adequate source of data to identify such patients. If this hypothesis holds true, then risk adjustment methods could also incorporate an indicator for NH care by relying on commonly used claims databases. To date, only 1 recent study has tested a relevant claims-based algorithm. This study by Zuckerman et al,¹⁰ using the MarketScan Medicare Supplemental and Coordination of Benefit Database, identified NH patients based on a combination of procedure codes, provider type, and place of service. This algorithm yielded a sensitivity of 96.7%, specificity of 99.9%, and a Kappa statistic of 0.97.

We test our hypothesis using a database developed by linking records from Ohio's cancer surveillance system with Medicare and MDS data, as described below.

Methods

Overview

To study our research question, we identified patients receiving NH care separately through the MDS file and through the Carrier files. We then examined a cross-tabulation of the data by the 2 sources and calculated the sensitivity and positive predictive value (PPV) of claims data by considering the MDS as the gold standard.

Sources of Data

Data were obtained from the Ohio Cancer-Aging Linked Database (CALD), which was developed by linking records from the Ohio Cancer Incidence Surveillance system (OCISS) with multiple data sources, including the Medicare claims and enrollment files, and the NH MDS. The CALD was developed as part of ongoing studies examining disparities in cancer-related outcomes in Ohio residents, 65 years of age or older, and diagnosed with incident breast, prostate, or colorectal cancer during the years 1997–2001.

The datasets were linked using patient identifiers, including social security number and sex. Records from the OCISS and Medicare files were linked by the Centers for Medicare and Medicaid Services (CMS), whereas the investigative team linked records from the OCISS and the MDS. At each step of the linkage, we appended the OCISS dummy patient identifier (PATID), and dropped the more sensitive identifiers from the analytic files. PATID was then used to identify patients across the various files. Approval for the development of the Ohio CALD and relevant studies was obtained from the Ohio Department of Health, which administers the OCISS, as well as CMS, and the Institutional Review Board of the University Hospitals of Cleveland.

The MDS includes assessment data for all patients admitted to a Medicare or Medicaid certified NH. Each record in the file represents an assessment, which is completed upon admission, and every 90 days thereafter. Therefore, residence in a NH in a given time period would be confirmed if the patient had at least 1 assessment completed during that period. In addition to rich clinical, functional, cognitive, and psychosocial assessment data, MDS records carry patient identifiers and the date at which the assessment was completed. This study relied only on the latter 2 variables to confirm patient NH status during the study period.

The Carrier Standard Analytic File (SAF) includes claim records submitted by noninstitutional providers, such as physicians.¹¹ This is different from the Outpatient SAF, which includes claims submitted by institutional outpatient providers, including hospital outpatient departments. In addition to dates of service and diagnosis codes, claim records from the Carrier SAF carry procedure codes in *Current Procedural Terminology, 4th Edition* (CPT-4), and Healthcare Common Procedure Coding System.

Finally, the denominator file was used to identify patients enrolled in managed care programs.

Patient Population

Using the MDS file, we identified all patients with any MDS assessment during calendar year 2002 (n = 5327). We excluded patients who were enrolled in managed care during the study year (n = 698), leaving the MDS cohort at 4629 patients.

From carrier claim records carrying service dates November 1, 2001 through December 31, 2002, we used procedure codes indicating E&M of a patient residing in a NH (CPT-4 99301–99303 and 99311–99313) to identify NH patients (n = 5058). A total of 199 patients were excluded because of participation in managed care programs, leaving the cohort of NH patients identified through the carrier file at 4859.

Analysis

We examined a cross-tabulation of data by the 2 sources and calculated sensitivity and PPV. Sensitivity reflects the percentage of patients in 2002 MDS file that were identified in 2002 and last 2 months of 2001 carrier file. PPV is defined as the percentage of patients with the relevant E&M codes in the carrier file that were found in the MDS file. Specificity is not applied because the number of the patients unidentified from both files is zero. All analyses were performed using SAS V9.1 (SAS institute Inc, Cary, NC).

Results

As shown in the Table 1, 5412 patients were identified either through the MDS or claims files as having received NH care in 2002. We report a sensitivity of 88.1%, as the proportion of patients identified through the MDS file who were also successfully identified as

recipients of NH care through the carrier file. The PPV, or the proportion of patients with relevant E&M codes in the Carrier file who were also identified in the MDS file, was 83.9%. Using the carrier files alone, we were unable to identify 10.2% of NH patients and 14.4% were mistakenly identified as being in a NH.

We conducted 2 types of sensitivity analyses. In the first set of analyses, we required at least 2 claims with the relevant procedure codes in the carrier file. This analysis yielded a considerably lower sensitivity (75.8%), but a slightly higher PPV (86.7%). Based on these results, we believe that requiring at least 2 claims in the carrier file to ascertain a patient's NH status would miss cases with short NH stays. This speculation was supported by the finding showing a smaller average number of MDS assessments among patients with 1 claim only versus those with at least 2 claims with the relevant procedure codes (2.7 vs. 5.2).

Next, we incorporated in our algorithm the additional procedure codes listed by Zuckerman et al, as well as the relevant codes for provider type and place of service. Our results changed only slightly and did not reach the sensitivity reported by Zuckerman et al.

Discussion

Our findings imply that the use of physician service CPT codes for NH care taken from the Carrier files have good sensitivity for detecting patients who reside in a NH. Such an approach would allow the identification of NH residents without resorting to MDS or to state Medicaid data. The simplicity of our algorithm, utilizing only 2 ranges of procedure codes, may make its use to identify NH patients more widespread among users of claims data with little loss of sensitivity compared with the more complex approach proposed by Zuckerman et al. The differences in our levels of sensitivity may emanate from differences in the data being used, with ours being especially relevant for those using Surveillance, Epidemiology, and End Results (SEER)-Medicare linked files.

The reasons that 10.2% of the patients in nursing homes are not identified through this approach may reflect a failure to bill for services on the part of physicians or an incorrect use of CPT codes for NH care. Probably, the most common reason that no bill is submitted is for patients with very short stays. Although regulations require an admission visit by a physician, in many settings physicians only visit the NH a few times a month and if the stay is short either because of death, readmission to the hospital, or discharge to the community there indeed may have been no encounter to bill.

Use of incorrect CPT codes presumably accounts for the remainder of the shortfall in carrier file sensitivity to NH residence. In this study we found that about 10% of patients who were in the NH had carrier bills indicating the care occurred in the NH but the CPT code used did not reflect such care. There are likely others where both codes are in error.

The 14.4% of patients with CPT codes indicating that they received service in a NH but who do not have MDS assessments on file is more difficult to explain. There are 2 settings in which these CPT codes are appropriate but where MDS assessments are not done: intermediate care facilities for persons with mental retardation and residential psychiatric facilities,¹² but neither of these are listed with significant frequency in the carrier file in this group of patients with no MDS assessments. A second cause for lack of MDS assessment is facilities that have had their Medicare or Medicaid certification removed as part of the inspection efforts, but this number is likely very small.¹³ The most likely explanation is that these bills are submitted for patients who are residing in mixed facilities and the billing clinician perceives the patient to be in a NH when in fact they might, for example, be in assisted living where MDS evaluations are not used. In 98% of the patients described here with E&M CPT codes for NH patients and no corresponding MDS assessment, at least one

of the CPT codes reported has a place of service listed as being in a NH. In this same group however there are on average 4 NH CPT codes reported with place of service not in a NH, supporting the notion that place of service codes are frequently inaccurate as suggested to us by the Research Data Assistance Center, an organization contracted by CMS to assist researchers in utilizing administrative data. It should be noted however that care of patients in NH often legitimately includes care in other places and in this population of patients with cancer this may be more frequent than in other groups. Further work with explicitly examining MedPAR and Medicaid claims of patients in this group might help to clarify the actual location of these patients whose NH residence is ambiguous.

To identify NH patients, one would ideally use MDS files. However, this approach may be impractical for studies underway, especially for the ones using databases such as the SEER-Medicare data, to which appending other data sources may not be feasible. When using state cancer surveillance data, obtaining MDS data requires considerable effort. The Medicare Carrier file may be a good surrogate for patient NH status. Using the claims-based algorithm presented in this study, future analyses should be directed toward the development of case-mix adjustment methodologies accounting for NH status.

Acknowledgments

Supported by a Career Development Grant from the National Cancer Institute grant K07 CA096705 (to S.M.K.), and a NIH Cancer-Aging Research Development Grant P20 CA103736 (to Dr. Nathan Berger, Case School of Medicine; S.M.K., pilot project investigator).

REFERENCES

1. Kasper, J.; O'Malley, M. [Accessed March 6, 2008] Changes in characteristics, needs, and payments for care of elderly nursing home residents: 1999 to 2004. The Kaiser J Family Foundation. June. 2007 Available at: <http://www.kff.org/medicaid/upload/7663.pdf>
2. American Health Care Association. [Accessed March 6, 2008] The state long-term health care sector 2005: characteristics, utilization, and government funding. August 29. 2006 Available at: http://www.ahca.org/research/statestatsrpt_20060823_final.pdf
3. Klabunde CN, Warren JL, Legler JM. Assessing comorbidity using claims data. An overview. *Med Care*. 2002; 40:IV-26-IV-35.
4. Klabunde CN, Potosky AL, Legler JM, et al. Development of a comorbidity index using physician claims data. *J Clin Epidemiol*. 2000; 53:1258-1267. [PubMed: 11146273]
5. Koroukian SM, Xu F, Beaird H, et al. Complexity of care needs and unstaged cancer in elders: a population-based study. *Cancer Detect Prev*. 2007; 31:199-206. [PubMed: 17658225]
6. SEER. [Accessed November 23, 2005] Glossary of Statistical Terms. Available at: <http://seer.cancer.gov/cgi-bin/glossary/glossary.pl#27>
7. Bradley CJ, Clement JP, Lin C. Absence of cancer diagnosis and treatment in elderly Medicaid-insured nursing home residents. *J Natl Cancer Inst*. 2008; 100:21-31. [PubMed: 18159068]
8. Bradley CJ, Given CW, Dahman B, et al. Diagnosis of advanced cancer among elderly Medicare and Medicaid patients. *Med Care*. 2007; 45:410-419. [PubMed: 17446827]
9. Bradley CJ, Given CW, Luo Z, et al. Medicaid, Medicare, and the Michigan tumor registry: a linkage strategy. *Med Decis Making*. 2007; 27:352-363. [PubMed: 17641138]
10. Zuckerman IH, Sato M, Hsu VD, et al. Validation of a method for identifying nursing home admissions using administrative claims. *BMC Health Serv Res*. 2007; 7:202. [PubMed: 18070360]
11. Medicare Data File Descriptions. Research Data Assistance Center (ResDAC); Available at: http://www.resdac.umn.edu/medicare/file_descriptions.asp#carrier
12. Centers for Medicare & Medicaid Services (A). [Accessed November 2007] Medicare Claims Processing Manual Chapter 12-Physicians/Nonphysician Practitioners. Available at: <http://www.cms.hhs.gov/manuals/downloads/clm104c12.pdf>

13. Centers for Medicare & Medicaid Services (B). [Accessed November 2007] Medicare Claims Processing Manual Chapter 6–SNF Inpatient Part A Billing. Available at: <http://www.cms.hhs.gov/manuals/downloads/clm104c06.pdf>

TABLE 1

Sensitivity and Positive Predictive Value of Patients Identified From Carrier File as Utilizing Nursing Home Care During the Period November 2001 through December 2002, Matched With Patients in Minimum Data Set Data During Calendar Year 2002

	Patients Identified From MDS Data			(Row) Percentage
	Yes	No	Total	
Patients identified from the carrier file, through E&M codes				
Yes	4076	783	4859	83.9 (PPV)
No	553	0	553	100.0
Total	4629	783	5412	
(Column) Percentage	88.1	100.0	89.8	
(Sensitivity)				