Risk Adjusting Survival Outcomes in Hospitals That Treat Patients With Cancer Without Information on Cancer Stage

David G. Pfister, MD; David M. Rubin, BS; Elena B. Elkin, PhD, MPA; Ushma S. Neill, PhD; Elaine Duck, RN, MA, MS; Mark Radzyner, JD, MBA; Peter B. Bach, MD, MAPP

IMPORTANCE Instituting widespread measurement of outcomes for cancer hospitals using administrative data is difficult owing to lack of cancer-specific information such as disease stage.

OBJECTIVE To evaluate the performance of hospitals that treat patients with cancer using Medicare data for outcome ascertainment and risk adjustment and to assess whether hospital rankings based on these measures are altered by the addition of cancer-specific information.

DESIGN, SETTING, AND PARTICIPANTS Risk-adjusted cumulative mortality rates of patients with cancer were captured in Medicare claims data from 2005 through 2009 nationally and assessed at the hospital level. Similar analyses were conducted using Surveillance, Epidemiology, and End Results (SEER)-Medicare data for the subset of the United States covered by the SEER program to determine whether the inclusion of cancer-specific information (only available in cancer registries) in risk adjustment altered measured hospital performance. Data were from 729,279 fee-for-service Medicare beneficiaries treated for cancer in 2006 at hospitals treating 10 or more patients with each of the following cancers, according to Medicare claims: lung, prostate, breast, colon, and other. An additional sample of 18,677 similar patients were included from the SEER-Medicare administrative data.

MAIN OUTCOMES AND MEASURES Risk-adjusted mortality overall and by cancer category, stratified by type of hospital; measures of correlation and agreement between hospital-level outcomes risk adjusted using Medicare data alone and Medicare data with SEER data.

RESULTS There were large survival differences between different types of hospitals that treat Medicare patients with cancer. At 1 year, mortality for patients treated by hospitals exempt from the Medicare prospective payment system was 10% lower than at community hospitals (18% vs 28%) across all cancers, and the pattern persisted through 5 years of follow-up and within specific cancer categories. Performance ranking of hospitals was consistent with or without SEER-Medicare disease stage information (weighted κ ≥ 0.81).

CONCLUSIONS AND RELEVANCE Potentially important outcome differences exist between different types of hospitals that treat patients with cancer after risk adjustment using information in Medicare administrative data. This type of risk adjustment may be adequate for evaluating hospital performance, since the additional adjustment for data available only in cancer registries does not seem to appreciably alter measures of performance.

Published online October 8, 2015. Corrected on November 12, 2015.

Copyright 2015 American Medical Association. All rights reserved.
Cancer is a leading cause of mortality. Decades of research have demonstrated that outcomes of cancer treatment vary widely in relation to where patients receive their care, and there are widespread concerns about cancer care costs. As a result, a number of initiatives are under way. The Center for Medicare & Medicaid Innovation has published its plans to bundle reimbursements in oncology. United HealthCare has tied payment changes in cancer to quality measurement in a pilot program that is now being expanded. Anthem has implemented programs where oncologists are paid a bonus for following some treatment approaches but not others.

Quality measures suitable for these initiatives, although rising in number, have shortcomings. In the list of 59 quality measures relevant to cancer endorsed by the National Quality Forum, most (76%) describe processes of care. Of these, a sizable fraction (13%) are purely retrospective, focusing on care received by patients prior to their death rather than on health outcomes. In the National Quality Forum database, at least two-thirds of quality measures related to cancer require chart review to ascertain disease stage or other detailed clinical information.

Although survival is perhaps the most important outcome to patients with cancer and can be readily ascertained from administrative data, researchers hesitate to rely on administratively derived data for risk adjustment owing to concerns that comparisons between hospitals will be confounded by underlying differences in treated populations. In large part, these concerns arise from the lack of potentially critical information about a patient's cancer, such as stage and timing of a cancer diagnosis, in administrative data. While this information predicts outcomes at the individual level, it is not known whether it would have a large influence on risk-adjusted performance at a more aggregate level, such as the level of the treating hospital. If hospitals varied systematically in the cancer stage distributions of their patient populations, it may; however, if true differences in performance are large, it may not.

Our objective was to evaluate risk-adjusted performance (as measured by survival) of different types of hospitals that treat patients with cancer using information from health insurance claims for risk adjustment and then to assess in a parallel data set whether the risk-adjusted performance of hospitals is robust to the inclusion of patient-level information on cancer stage and date of diagnosis.

Methods

Data Sources

We analyzed 2 parallel data sets: (1) a national data set of fee-for-service Medicare claims from across the United States, and (2) the Surveillance, Epidemiology, and End Results (SEER)-Medicare database, which links the National Cancer Institute (NCI)-sponsored consortium of population-based cancer registries linked to the Medicare claims and enrollment information that covers almost 26% of the US population. The 2 analyses were run in parallel using identical methods, and we assumed that findings in the SEER-Medicare data regarding the robustness of hospital-level evaluation are generalizable to the analyses of the nationwide Medicare program.

Cohort Selection and Hospital Assignment

Both analytic cohorts were made up of individuals who appeared to be beginning cancer treatment or beginning management of recurrent disease in 2006, indicated by an absence of claims for cancer in 2005 and either inpatient or outpatient claims with a primary or secondary diagnosis code for cancer on a claim for a cancer-related service in 2006 or a primary diagnosis code of cancer on a claim for a service not clearly cancer related in 2006. Cancer services were identified by the Healthcare Common Procedure Coding System. The type of cancer was determined from hierarchical algorithms evaluating listed diagnoses that prioritized inpatient claims, then outpatient hospital claims, then physician claims, where the predominant diagnosis was selected within each category. The type of cancer identified by this algorithm was nearly always identical to the cancer diagnosis recorded in SEER (eTable 1 in the Supplement) in that analysis. We limited our analysis to patients with only 1 type of cancer listed.

Patients who were not continuously enrolled in Part A and B Medicare in 2005 were excluded, as were those not continuously enrolled from 2006 to either their death or December 31, 2009, whichever came first. Patients were assigned to a single hospital based on claims within the 180 days following the first claim for cancer treatment in 2006. We used a hierarchical approach to assignment (eTable 2 in the Supplement). Primary assignment to a hospital that delivered some portion of the patient’s care was possible in 89% of patients; 8% were assigned to the hospital where their physician accrued the greatest amount of Medicare reimbursement; 3% were assigned through physicians who shared patients with the study patient’s physician.

For the SEER-Medicare analyses, we eliminated hospitals (and those patients assigned to them) if they had fewer than 10 patients treated for any of the 4 major cancer diagnoses (lung, breast, colorectal, or prostate), yielding a study cohort of 18,677 patients. We then searched the noncancer file, which is a representative subset of patients without a cancer diagnosis recorded in the SEER registry, for patients with a Medicare claim containing a diagnosis code for cancer. We found...
very few such patients (for 94% of the hospitals, the count of additional patients found was 1 or 0); therefore, such a small number would not have influenced our findings. In addition, the essence of our analysis was to incorporate cancer-specific data, which these patients by definition lacked because they had no record of cancer diagnosis in SEER.

Risk Adjustment

Analyses were risk adjusted based on information available in claims using the 3M Clinical Risk Group (CRG) software. The CRG classification system is used by payers and health authorities for risk adjustment in quality reporting, rate setting, and utilization review. The software has been used by Medicare and various state Medicaid programs for purposes ranging from quality reporting to payment policy.16,17 The CRG algorithm assigns individuals to 1 of 1080 mutually exclusive groups that reflect overall health status and the presence and severity of specific diseases and health conditions using diagnostic and procedure codes and beneficiary age and sex. In addition to this risk adjustment we also controlled for median household income in the zip code of residence, classified in tertiles of the national household income distribution from US census data. We empirically selected the period of time we would assess to capture risk-adjustment data, settling on a 180-day period after the first cancer treatment claim, a stopping point that assured that 99% of patients sampled were no longer assigned CRG status 1 (healthy).

In the SEER-Medicare data, we additionally stratified for the stage of the person’s cancer, and whether their date of cancer diagnosis was around the time of their initial claim for treatment. Cancer stage was divided into 5 categories, stages I, II, III, IV, and unknown, based on the American Joint Committee on Cancer stage classification.18 We created a binary indicator to categorize cancers as incident if the date of diagnosis occurred within 4 months prior to the first claim we found for cancer treatment in 2006; otherwise, they were categorized as prevalent.

Analysis

Expected and observed overall survival were compared from the date of the first cancer treatment claim in 2006 to end of follow-up (December 31, 2009) or the date of death from any cause. In Medicare analyses we compared the survival between mutually exclusive categories of hospital type: (1) free-standing cancer hospitals that are exempt from the Medicare prospective payment system (PPS) (n = 11); (2) the remaining NCI-designated cancer centers that had adequate numbers of patients (n = 32); (3) other academic teaching hospitals (n = 252); and (4) remaining hospitals, labeled “community hospitals” (n = 4873).

In SEER-Medicare analyses we compared observed to expected overall survival outcomes that were Medicare risk-adjusted with those that were SEER-Medicare risk adjusted. The expected rates of survival were determined for each individual patient based on the average for similar patients stratified by cancer type, CRG category, sociodemographic characteristics, age at the time of initial treatment (66-69, 70-79, and ≥80 years), sex, and median income tertile. Individual level and grouped hospital performance was then evaluated by taking the quotient of the sum of the actual deaths and the sum of the expected deaths for each hospital or hospital category.

To assess the impact of cancer-specific information on risk-adjusted hospital performance, we evaluated the correlation of hospital rankings within the SEER-Medicare data between their Medicare and SEER-Medicare risk-adjusted outcomes divided into quintiles of survival at 3 years and 5 years. We then assessed the stability of ranking by determining the proportion of hospitals that moved either 1 or 2 quintiles in rank between the 2 risk adjustment approaches, and measured overall agreement in ranks using the Cohen weighted κ statistic, with weights proportional to the square of the distance from the diagonal (ie, quadratic weighting). Kappa values range from 0 to 1, with higher values reflecting higher correlation.19,20 Authors disagree on exactly how high κ values need to be to reflect acceptable correlation. Landis and Koch21 argue that values of 0.81 or higher constitute near perfect agreement; Fleiss and colleagues22 characterized values greater than 0.75 as excellent.23,24 The proportions of between-hospital variation are explained, and the linear correlations of rankings are displayed in the Supplement. We compared rankings overall and within each of 4 common cancer types and in an “other” category to parallel our analyses of performance by different cancer hospital types.

The study was deemed exempt research by the institutional review board at Memorial Sloan Kettering Cancer Center, and the SEER-Medicare files were used in accordance with a data use agreement from the NCI.

Results

Medicare Analysis of Different Types of Cancer Hospitals

Figure 1 shows the sample size and the outcomes of risk-adjusted death over 5 years across the cancer diagnoses included in the analysis for each of the 4 major cancer types and for the other cancer category, both cumulatively and annually, according to the type of hospital providing cancer care. In general, both the annual probability of death and the cumulative probabilities of death aligned with the category of hospital in a direction consistent with the findings of prior outcome studies, both within each cancer type and overall.4,25,26 The risk-adjusted probability of death at 1 year for all patients under the care of PPS-exempt cancer hospitals was 10% lower than that at community hospitals (18% vs 28%), with other types of hospitals falling between the 2 extremes. Successive years followed a similar pattern, with the overall survival gap persisting and the conditional probabilities of death in each year being lowest among patients cared for at the PPS-exempt hospitals and highest in community hospitals. A similar pattern was seen in each of the cancer-specific analyses.

Medicare Risk-Adjusted vs SEER-Medicare
Risk-Adjusted Outcomes

For these analyses, 18 677 patients met eligibility criteria, received treatment for cancer in 2006, had no claim for care as-
associated with cancer in 2005, could be assigned to a hospital or physician, and had known cancer stage. Of them, 36% had cancer of the lung, 33% prostate, 23% breast, 6% colorectal, and 2% other type (Table 1).

When comparing risk-adjusted hospital performance for these patients, we found that for each cancer, and for all cancers combined, correlation in quintiles of rank between Medicare-adjusted and SEER-Medicare-adjusted outcomes had very high κ values for 3-year and 5-year survival (Table 2). All exceeded the 0.81 cutoff proposed by Landis and Koch for nearly perfect correlation. Consistent with this finding, empirical shifts in rank of hospitals under the alternative methods of risk adjustment were uncommon (Figure 2). For instance, with respect to 5-year survival, overall, only 3% of hospitals moved 2 quintiles or more; 11% moved by at least 1 quintile. In the 4 major cancer types, in no case did more than 2% of hospitals...
move 2 or more quintiles in ranking, and between 1% and 15% moved at least 1 quintile.

Other metrics, such as the extent of between-hospital variation and correlation of performance measures, also showed agreement (eFigure 1 and eFigure 2 in the Supplement). For instance, for all cancers combined, there was an unadjusted gap between 25th and 75th percentile survival probabilities of 24%. Risk adjustment led to reductions in the magnitude of variation to a similar degree with either approach, with 25th to 75th percentile gaps of 10% after Medicare risk adjustment and 8% after SEER-Medicare risk adjustment. In terms of absolute correlation, all $R^2$ values were 0.80 or greater.

Sensitivity Analysis

Our findings did not change meaningfully when we excluded from analysis the 11% of patients who were assigned to a hospital indirectly in the SEER-Medicare data either through association with a physician who treated other patients in the assigned hospital or through the admitting patterns of that physician with other patients.

Discussion

A recent report from the Institute of Medicine27 raises far-reaching questions about the quality of cancer care available to patients. The scrutiny is appropriate. Measures such as hospital-specific volume for particular types of cancer care have regularly been associated with variations in both short- and long-term outcomes, as have other umbrella designations such as hospital teaching status or NCI designation.2,4,26 In these prior analyses, risk adjustment generally included cancer-specific data on stage and timing of diagnosis, and so the geographic scope of these analyses is limited to the few parts of the country where such information has been routinely linked to claims. The SEER-Medicare data set is the most widely used resource for these types of studies, and it currently covers about 26% of the US population.
In our analysis, we also show large and persistent risk-adjusted differences in cancer treatment outcomes associated with the type of treating hospital. The findings suggest that compared with community hospitals, survival appears to be superior for patients treated at PPS-exempt cancer hospitals, at NCI-designated cancer centers, and at academic teaching hospitals—all findings consistent with prior reports. But because these analyses only use Medicare data for risk adjustment, a critical question is whether the lack of cancer-specific data on each treated patient, which can only be obtained readily from cancer registries, matters for performance assessment at the hospital level.

Our findings examining SEER-Medicare data strongly suggest that the disease-specific information available in cancer registries, although undoubtedly influential on individual patient outcomes, may not be routinely needed for risk adjustment of performance measures at the hospital level. Comparing outcomes adjusted for only Medicare data and then with both Medicare and SEER data, we found that hospital ranks were stable; weighted κ values signaled very...
high agreement; the explanatory power for overall variation was similar; and linear measures of performance scores were highly correlated.

In other words, using Medicare claims alone to identify, risk adjust, and evaluate outcomes may be an adequate way of understanding cancer care at the hospital level. If true, this would align cancer care outcome evaluation with many measures for other conditions that rely on administrative data alone and have been endorsed by the National Quality Forum and incorporated by the Centers for Medicare & Medicaid Services (CMS) in programs for assessing quality and calculating value-based payment rates. Some measures cover cardiovascular disease outcomes; others in development focus on orthopedic care and pneumonia. Each relies on administrative claims to both capture events and to risk adjust. Co-morbidities and other types of risk-adjustment are intrinsic controversy. Perhaps the most common rebuke to any outcome measure is the concern that patients are sicker than the method could capture. The comorbidity grouping technique used in the present study relies on the CRG commercial software from 3M that is designed for gauging outcomes and forecasting expenditures, the purposes for which we are using it here. Although it is proprietary, its highly detailed documentation is in the public domain, and the software is already in use by the Medicare program and several state Medicaid programs for various purposes.

Performance measurement and ranking itself are also subjects of controversy. Critics of the approach might argue that even a reclassification of a handful of hospitals under alternative risk-adjustment approaches undermines the credibility of the endeavor. We believe, instead, that our findings support cautious movement toward profiling hospitals for their performance and perhaps using those assessments for quality-assessment and payment initiatives. We believe this because we gauge the impact on patients of the outcome differences we report to be of much more importance than the possible small impact of occasional misclassification of a hospital, particularly since the findings of our analyses showing hospital differences involve avoidable deaths of patients, while the anticipated impact of hospital misclassification would be modest and economic. A focus on risk adjusted outcomes would also direct more attention to care processes that might affect outcomes even if they are not currently measured or subject to public reporting.

In our analyses we include factors that adjust for a patient’s socioeconomic status while CMS risk adjustment in some of its programs does not. This divergence from CMS methods reflects a philosophical difference of opinion. We believe that economic status is associated with outcomes and thus should be considered. We also capture comorbidities after diagnosis, while other approaches only look to the year prior to the hospitalization for risk adjustment. We take this extra step to assign CRGs to all patients with an expectation that comorbidities and other cancer-specific factors associated with outcome are likely ascertained in a more accurate fashion during the process of initial management of a patient’s cancer rather than before their diagnosis.

Our findings suggest an opportunity to use administrative data to assess the quality of cancer care provided by US hospitals. The major impediment to doing so has been a long-held belief that adequate risk-adjustment could not be accomplished without information regarding stage of disease and other details included in cancer registries and medical records. We find no support for that belief. Rather, information on cancer stage and timing of diagnosis adds little to insights garnered from administrative data on hospital performance.

That there are very sizable differences in outcomes between hospitals may explain why stage data seem not to be important—the actual differences are of much greater magnitude than small differences stage mix could explain. These large differences reinforce the conclusion drawn by the Institute of Medicine that the quality of cancer care in the United States is inconsistent and should be improved. To do so, we must first be able to observe, measure, and compare it both reliably and efficiently. The methodology we describe here provides a possible starting point.
Author Contributions: Mr Rubin had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Pfister, Rubin, Neill, Duck, Radzyner, Bach.

Acquisition, analysis, or interpretation of data: Pfister, Rubin, Elkin, Neill, Duck, Radzyner, Bach.

Drafting of the manuscript: Pfister, Rubin, Elkin, Neill, Duck, Radzyner, Bach.

Critical revision of the manuscript for important intellectual content: Pfister, Rubin, Elkin, Neill, Duck, Radzyner, Bach.

Statistical analysis: Rubin, Duck, Radzyner.

Administrative, technical, or material support: Pfister, Rubin, Neill, Duck, Radzyner, Bach.

Study supervision: Pfister, Radzyner, Bach.

Conflict of Interest Disclosures: All authors are employed by a PPS-exempt cancer hospital, Memorial Sloan Kettering Cancer Center. No other conflicts are reported.

Funding/Support: This study was funded by internal Memorial Sloan Kettering Cancer Center funds and by Memorial Sloan Kettering Cancer Center Support Grant/Core Grant P30 CA008748.

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review or approval of the manuscript; decision to submit the manuscript for publication.

Additional Contributions: We would like to thank Geoffrey Schnorr, BS (Memorial Sloan Kettering Cancer Center), for research assistance, and Norbert Goldfield, MD (3M), for his insight and guidance with respect to the 3M CRG software, neither of whom were compensated for their contributions.

Correction: This article was corrected on November 12, 2015, to fix erroneous data reports in Table 2.

REFERENCES