Published in final edited form as:

Cancer. 2011 March 1; 117(5): 1038–1048. doi:10.1002/cncr.25677.

Using Claims-Based Measures to Predict Performance Status Score in Lung Cancer Patients

Ramzi G Salloum, MA¹, Thomas J Smith, MD², Gail A Jensen, PhD³, and Jennifer Elston Lafata, PhD^{1,4}

¹Center for Health Services Research, Henry Ford Health System, Detroit, Michigan

²Division of Hematology/Oncology and Palliative Care and Massey Cancer Center, School of Medicine, Virginia Commonwealth University, Richmond, Virginia

³Institute of Gerontology and Department of Economics, Wayne State University, Detroit, Michigan

⁴Social and Behavioral Health and Massey Cancer Center, School of Medicine, Virginia Commonwealth University, Richmond, Virginia

Abstract

Background—Performance status (PS) is a good prognostic factor in lung cancer and used to assess chemotherapy appropriateness. Researchers studying chemotherapy use are often hindered by the unavailability of PS in automated data sources. To our knowledge, no attempts have been made to estimate PS using claims-based measures. This study explores the ability to estimate PS using routinely available claims-based measures.

Methods—A cohort of insured patients aged 50+, diagnosed with stage II–IV lung cancer between 2000–2007 was identified via a tumor registry (n=552). PS was abstracted from medical records. Automated medical and pharmaceutical claims from the year preceding diagnosis were linked to tumor registry data. A logistic regression model was fit to estimate good vs. poor PS in a random half of the sample. C statistics, sensitivity, specificity, and R^2 were used to compare the predictive ability of models that included demographic factors, comorbidity measures, and claims-based utilization variables. Model fit was evaluated in the other half of the sample.

Results—PS was available in 80% of medical records. The multivariable regression model predicted good PS with high sensitivity (0.88/0.94 depending upon how good PS was defined), but moderate specificity (0.45/0.32) with a 0.50 prediction cutoff, and good sensitivity (0.64/0.83) and specificity (0.69/0.55) when the cutoff is 0.70. Goodness-of-fit c-statistic was 0.76/0.78.

Conclusions—PS can be estimated, with some accuracy, using claims-based measures. Emphasis should be placed on documenting PS in medical records and tumor registries.

Keywords

performance status; lung cancer; automated claims; medical chart; tumor registry; chemotherapy

Since 1997 evidence-based guidelines have recommended the use of chemotherapy treatment for medically fit patients with lung cancer to improve survival, symptoms, and quality of life.^{1–4} Despite these recommendations, numerous studies^{5–8} have illustrated variability in receipt of chemotherapy treatment among patients with lung cancer. Yet, the

ability to determine the appropriateness of observed treatment variability has been greatly hindered by voids in the clinical information necessary to judge appropriateness.

One key factor in evaluating the appropriateness of chemotherapy treatment is the patient's performance status (PS).^{1–4} Performance status is a subjective composite measure used by clinicians to assess current functional capacity, and the likelihood of adverse events, quality-of-life, and survival after treatment. Measures of PS are currently not available through automated medical claims, tumor registries, or other observational data commonly used to study cancer treatment and its associated outcomes. Thus, use of such data to address questions regarding chemotherapy treatment has been relatively limited and when undertaken the inability to consider PS is a noted limitation.^{8–10} The systematic lack of information on PS similarly impedes the ability of researchers to use existing automated, observational data for comparative effectiveness research.

This article asks two questions. First, how often are measures of a patient's PS documented in his or her detailed medical record? Second, is it possible to accurately estimate a patient's PS using routinely available tumor registry and claims-based measures on that patient's demographics, comorbidities, and prior healthcare utilization? Using a cohort of lung cancer patients diagnosed between 2000 and 2007, the feasibility of using medical record documentation to obtain PS measures is described overall and by patient characteristics. We then combine medical record documented PS information with information routinely available in an automated tumor registry as well as medical and pharmaceutical claims data to evaluate the feasibility of estimating PS among lung cancer patients using information routinely available in observational data sources. To our knowledge, this has not previously been attempted among patients with lung, or other, cancers.

METHODS

Study Population and Setting

Study patients were those receiving care from a 900-physician member, multi-specialty, salaried medical group practice in southeast Michigan. Data available from the medical group's tumor registry were used to identify all patients aged 50 and older who were diagnosed with lung cancer between January 1, 2000 and December 31, 2007. The medical group, which provides care under both fee-for-service and capitated arrangements, staffs 27 primary care clinics throughout Detroit and the surrounding metropolitan area. Patients eligible for study inclusion were those continuously enrolled in an affiliated health plan (i.e., health maintenance organization) for the one-year period preceding their date of lung cancer diagnosis. Patients for whom no stage was available at the time of diagnosis or for whom the stage at diagnosis was 0-I were excluded, as chemotherapy treatment was not indicated for patients staged 0 or I during this time period.¹¹ The Medical Group's Institutional Review Board (IRB) approved all aspects of the study protocol.

Performance Status Measurement

The two most commonly used PS systems are the Eastern Cooperative Oncology Group (ECOG) scale, and the Karnofsky Performance Scale (KPS).¹² Although the two scales are not identical they are generally thought to capture the same conceptual domain and conversions are possible between them (Table 1).¹³

Two trained chart abstractors reviewed inpatient and outpatient nursing and physician notes available within the patient's electronic medical record from two months prior to diagnosis until the first of death, disenrollment, initiation of chemotherapy, or six months following diagnosis. If available, abstractors documented specific numeric performance scores and scale (i.e., ECOG or KPS). Patients were assigned a 'good' PS if they had an ECOG score of

0 or 1 or a KPS score of 100-80. A “poor” PS was assigned to patients with an ECOG score of 2–5 or a KPS score of 70-0. This was done to be consistent with standards in practice regarding recommendations for chemotherapy use among lung cancer patients during the study period^{1–4} as well as with existing research applications.¹⁴ With the issuance of the 2009 ASCO guidelines, the standard for chemotherapy use changed to include the consideration of use for those with ECOG score of 2 or a KPS score of 60–70. Thus, we also present alternative results for which those with these scores are realigned to PS ‘good.’

If no numeric score was documented, abstractors collected medical record documentation of ‘good’ or ‘poor’ PS. If no reference to PS was documented in the medical record, notes regarding the patient’s functionality (e.g., references to shortness of breath, use of wheelchair or other personal mobility devices, labor force participation, exercising habits, activities of daily living or other references to mobility) were recorded and used to estimate PS. Interrater reliability between the 2 abstractors was assessed on a random subset of 40 observations. The resulting Cohen’s Kappa was 0.88. Among the inter-rater reliability subset (N=40), each incident where the abstracted PS did not match between the two abstractors (N=3), one abstractor indicated ‘good’ or ‘poor’ while the other selected ‘unknown’ PS. For the final analytical database, these differences were reconciled by choosing ‘good’/‘poor’ over ‘unknown’.

Automated Tumor Registry and Claims Data

Automated tumor registry and claims data were used to obtain patient demographic characteristics, cancer stage, and diagnoses for each patient. Demographic measures included age, gender, and race. The age of the patient (in years) was recorded as of lung cancer diagnosis date. Clinical variables examined included stage at diagnosis and the Charlson comorbidity index.¹⁵ Cancer stage was reported using the American Joint Committee on Cancer (AJCC) stages II–IV. A dichotomous variable was created to control for AJCC stage IV patients in the regression. The Deyo adaptation of the Charlson Comorbidity Index and each of its component diagnostic subgroups were constructed using inpatient and outpatient diagnostic information available in the 12-month period preceding diagnosis.¹⁶ In addition, claims data provided information on prescription drugs dispensed, and medical care utilization in the 12-month period preceding lung cancer diagnosis.

Medical care utilization measures included those reflective of inpatient stays in a short stay hospital or skilled nursing facility (SNF), ambulatory care visits, emergency department visits, and utilization of home health services, same day surgery, and durable medical equipment (DME). For each person, inpatient utilization measures included the total number of distinct inpatient stays, the total number of inpatient days, and the average of length of an inpatient stay, for those with a nonzero number of stays. The number of outpatient visits was recorded, and in the regression a dichotomous variable was created to control for patients with nonzero outpatient visits. Similar dichotomous variables were constructed to reflect any drug dispensing and any DME utilization. The emergency department, home health, and same day surgery utilization variables measured the counts of visits incurred. We also evaluated the use of a count of the distinct number of medications dispensed during the baseline year, as recommended by Schneeweiss and colleagues.¹⁷ For this measure, medications whose first eight digits of the American Hospital Formulary Services code were equal were considered to be the same drug.¹⁸

Statistical Methods

Among the cohort of lung cancer patients, we report the frequency of documented PS in medical records, and describe the different ways PS was recorded. Systematic differences between patients for whom PS was recorded and patients for whom it was not recorded were

examined using two sample t-tests (or Wilcoxon rank sum tests) and chi-squared tests, depending on the nature of the characteristic. Similar analyses were conducted to compare unadjusted differences in patient characteristics by ‘good’ PS vs. ‘poor’ PS. Multivariable logistical regression models were fit to evaluate the feasibility of using routinely available observational data to predict ‘good’ vs. ‘poor’ PS. Three separate models were estimated, reflective of three different levels of comprehensiveness of observational data routinely available. The first regression model included only those variables typically available via tumor registries (demographics and stage). The second model included those same variables, plus measures of medical care utilization and diagnoses available in medical claims data. The third model added measures of prescription drug use routinely available via pharmaceutical claims.

For each model, a split sample cross-validation was used to check for model overfitting. C statistics, sensitivity, specificity, and R^2 were used to assess and compare the predictive ability of the different models. Initially, all variables were considered for inclusion. However, the final model in each of the three categories was fit using the stepwise-elimination method. Pairwise interactions were tested but did not enhance model prediction. Likewise, we evaluated the need to account for the non-independence of patients seen by the same physician, but as the intra-class correlation coefficient (ICC) was negligible (ICC = 0.01), we elected not to do so as not doing so enabled us access to additional assessment of model fit. The final models were estimated on the full sample and bootstrapping was used to replicate each final model 1000 times to create 95% confidence intervals around the c and R^2 statistics.¹⁹

To examine model discrimination, patients were ranked by their predicted probability of ‘good’ PS based on each model. Patients were then divided into deciles based on increasing predicted probability of ‘good’ PS and actual ‘good’ PS rates were reported among patients in all deciles to suggest how well models separated patients with ‘good’ PS from those with ‘poor’ PS.²⁰

We used SAS software, version 9.1.3 (SAS Institute Inc, Cary, North Carolina) for all analyses, and we considered $p < .05$ to be statistically significant.

RESULTS

Cohort Characteristics

Five hundred and fifty two patients met the criteria for study eligibility. The mean age at diagnosis was 67.4 (standard deviation [SD], 9.1). Of the study-eligible patients, 42 percent were female, while the race distribution was 69 percent white and 31 percent black. The AJCC staging distribution was as follows: 9 percent were stage II, 20 percent stage IIIA, 19 percent stage IIIB, and 52 percent stage IV. The average Charlson comorbidity score across the eligible sample was 2.8 (SD, 3.4), while the average number of distinct prescription drugs used in the year prior to diagnosis was 9.3 (SD, 7.1).

The average number of inpatient days in the year preceding diagnosis for the cohort—including those with no inpatient stays—was 2.9 days (SD, 7.5), while the average number of inpatient stays was 0.5 (SD, 0.8), resulting in an average inpatient length of stay of 5.0 days (SD, 5.2). The average number of outpatient visits was 5.7 visits (SD, 8.5) and the average number of emergency department visits was 0.6 (SD, 1.1) for the same time period. Across the study-eligible sample, 28 percent recorded any home health utilization, 3 percent had same day surgery, 12 percent incurred a DME dispensing, and 4 percent incurred a stay in a rehabilitation or skilled nursing facility. None incurred a hospice stay in the period preceding lung cancer diagnosis.

Medical Record Documentation of Performance Status

Of the 552 study eligible patients, PS was recorded in the medical record for 261 cases (47 percent). Among these, a numeric score was documented in 248 cases (95 percent), with the ECOG scale most often used (74 percent). For the remaining 13 patients, although a numeric score was not documented, explicit documentation was found of either 'good' or 'poor' performance score.

Among the 291 (53%) patients for whom PS was not recorded, there were 181 for whom there was a sufficient verbal description of the patient's functioning in either the physicians' notes, nurses' notes, or a combination of both to enable a determination of either a 'good' or 'poor' PS score. Thus, overall there were 442 patients (80%) for whom PS was determinable in their medical record.

Differences in patient characteristics by PS documentation level are reported in Table 2. The first two columns compare those patients for whom medical record documentation could be used to determine PS (known PS) with those for whom medical record documentation was insufficient to determine PS ('unknown' PS). As shown, patients with 'unknown' PS (n=110) did not differ significantly from those with a known PS (n=442) in terms of demographic or clinical characteristics or measures of medical care utilization.

Among patients with a known PS, the third and fourth columns of Table 2 compare patient characteristics between those who had a documented PS – either numeric or verbal with those for whom a PS was extrapolated based on notes in the medical record. No significant differences were observed for most measures. However, there were significant differences by gender, diagnosis of atherosclerotic cardiovascular disease, and the average number of inpatient days.

Patient Factors Associated with Performance Status

Among the 442 patients for whom PS is known, 290 patients (66%) had a 'good' PS using the pre-2009 definition of 'good' and 152 (34%) had a 'poor' PS. This changes to 76% 'good' and 24% 'poor' when those with a document numeric performance score of 2 are considered 'good' as would be the consistent with that in the 2009 ASCO guidelines for the use of chemotherapy. The unadjusted differences in patient characteristics by PS are illustrated in Table 3. Compared to patients with 'good' PS, patients with 'poor' PS were significantly older (69.7 vs. 66.4 years), more likely to be male (66% vs. 54%), have stage IV disease (64% vs. 44%), and have a significantly higher Charlson comorbidity score (3.6 vs. 2.4). Consistent with the latter finding, patients with 'poor' PS were significantly more likely to have been diagnosed with several of the individual components of the comorbidity index when compared to those with 'good' PS. Patients with 'poor' PS also incurred significantly more inpatient days (5.5 vs. 1.7) as well as longer lengths of stay (6.8 vs. 5.4 days) in the year prior to diagnosis, and were more likely to have incurred any outpatient visit, home health utilization, or DME use in the year prior to diagnosis. Also of note is that patients with 'poor' PS were significantly less likely to have undergone chemotherapy in the year following diagnosis (42% vs. 82%) (data not shown). Similar differences between the groups are found when those with PS =2 are realigned with the 'good' PS group, with two exceptions: statistically significant differences in gender and the prevalence of peripheral vascular disease no longer exist.

Predicting Performance Status

Results from the logistical regression models predicting 'good' vs. 'poor' PS defined the two ways are presented in Table 4. Results are presented for models fit on the full sample and include only significant ($p < 0.05$) variables per the stepwise regression. In the model

which included only tumor registry variables, only age at diagnosis and AJCC stage were selected. Diagnosis of chronic pulmonary disease, the number of inpatient stays, any outpatient visits, and the number of emergency department admissions were all added when information from medical claims data were considered. One more variable, the number of distinct prescription drugs, was added when information from pharmaceutical claims data was considered (model 3).

Statistical performance improved with the inclusion of additional explanatory variables (Table 4). Cross-validated c and R^2 values were never more than 0.01 smaller than fitted values. Using a predictive threshold of 0.50, high sensitivity (0.88/0.94, depending upon how 'good' PS is defined) was obtained with the best model (Model 3) but moderate specificity (0.45/0.32). Increasing the predictive threshold to 0.70 continued to yield relatively high sensitivity (0.64/0.83) and more moderate specificity (0.69/0.55) regardless of how 'good' performance score is defined.

Table 5 shows the actual and predicted 'good' PS rates for patients within each of the 10 deciles. As measured by the Hosmer-Lemeshow chi-square statistic,²⁰ all models have good calibration, where actual and predicted rates within each of the 10 deciles were not significantly different ($p=0.69$, $p=0.32$, $p=0.13$, for models 1–3 respectively) when PS = 2 is defined as 'poor' and likewise not significantly different ($p=0.92$, $p=0.63$, $p=0.98$, for models 1–3 respectively) when PS = 2 is defined as 'good.' Model discrimination was also improved with the inclusion of more explanatory variables.

DISCUSSION

Among a contemporary cohort of patients with stage II–IV lung cancer, we found explicit medical record documentation of PS less than half the time (47%). Review of nursing and physician notes led to PS to be determinable via medical records 80% of the time. Given the central role that PS plays in clinical decision making among lung cancer patients the lack of consistent medical record documentation is troubling. When documented, we found the distribution of performance status among the cohort—34% 'poor' PS (when those with PS = 2 are considered 'poor')—to be identical to the 34% with 'poor' performance status reported by Lilenbaum et al in contemporary clinical studies.¹⁴

We found that 'poor' PS among lung cancer patients with stage III–IV disease can be predicted reasonably well regardless of whether PS = 2 is considered 'good' or 'poor.' Furthermore, this was true regardless of the level of comprehensiveness of data used, but particularly for models that used information routinely available in medical claims data or medical and pharmaceutical claims data combined where c -statistics were all above 0.70. While the inclusion of information routinely available in medical claims data marginally improved model fit and predictive accuracy when compared to a model fit using only data available in tumor registries, the inclusion of information from pharmaceutical claims data did not substantively alter model fit, regardless of how 'good' PS is defined.

To our knowledge this is the first study to use observational data to estimate PS for lung, or any other, cancer patients. As such, these findings represent an important contribution to the field. These findings are important to our ability to monitor quality of care and appropriateness of chemotherapy, and to our ability to prospectively identify patients who may be appropriate (but not targeted) for clinical trial or palliative care/hospice enrollment without relying on expensive and time consuming primary data collection methods. Predictive models such as the ones presented here that rely on data routinely available within large, observational databases can also be used to augment comparative effectiveness research, including comparisons of different chemotherapy regimens as well as receipt of

chemotherapy vs. non-chemotherapy treatment and thereby greatly enhance the capabilities of existing electronic databases such as that available via SEER-Medicare data.

While our findings of significant differences in chemotherapy receipt by ‘good’ vs. ‘poor’ PS add face validity to the accuracy of the PS score abstracted from the medical record, the fact that 42% of patients with medical record documented ‘poor’ PS received chemotherapy in the year following diagnosis highlights the importance of attempts such as ours to make documented PS or PS proxies more readily available to those who monitor and study cancer care quality and outcomes. At the time of this study, national clinical practice guidelines for patients with non-small cell lung cancer unequivocally recommended chemotherapy for patients with PS 0 or 1.¹⁻³ These guidelines suggested that chemotherapy might “possibly” be of benefit in patients with PS 2, noting that those patients had been excluded from clinical trials. This was in line with expert opinions of the time²¹. More recent data have shown survival and quality of life benefits for PS 2 patients, although less than with ‘good’ PS, and the most recent ASCO guidelines are more supportive of chemotherapy treatment for patients with PS 2.²² Routine lung cancer chemotherapy among patients with PS ≥ 3 continues to not be recommended by any national professional organization. Chemotherapy use in patients with little chance of benefit and more chance of toxicity may delay discussion about prognosis and dying,²³⁻²⁶ which may lead to further poor quality of care such as the inappropriate use of mechanical ventilation or delays in referral to hospice, worse surviving caregiver quality of life, and high end of life care costs.²⁴ Without PS proxies, little can be done to use automated data sources to monitor and measure either the under- or over-use of chemotherapy receipt and its implications on patient and economic outcomes.

Our results should be interpreted in light of the following limitations. First, subjectivity is present in the assignment of PS. Even when assessed by a healthcare professional, PS scales are subjective in nature²⁵ and when estimated by physicians known to be prone to error²⁶, usually over-estimated.¹⁴ Thus, even if our model were 100% accurate, caution would have to be used in interpreting results dependent upon an accurate classification of PS. Nonetheless, the ability to develop a useful proxy measure of PS from existing observational data will help in the use of existing national data resources such as that available with SEER-Medicare data for comparative effectiveness research. Second, our models were developed on a relatively small sample and one that is specific to one delivery system. Thus not only should care be taken when generalizing findings, but our parsimonious models may exclude important predictors of PS available in observational data. Finally, identifying patients with ‘poor’ PS by their diagnoses and use of care via claims data poses its own limitations. For instance, DME use varies significantly based on differing personal preferences and practices in addition to restrictions on reimbursement by public and private insurers. Although claims for DME offer useful information, they identify only selected people with potentially disabling conditions.²⁷ The same is true of medical diagnoses—many of which are known to be under-captured in medical claims data—and prescription drug dispensing which reflect only those medications prescribed by physicians which the patient elected to fill. Yet, the ability to proxy PS is critical to the ability to use observational data to accurately draw conclusions about comparative effectiveness and cancer care quality at a population level if not at the bedside.

Despite these limitations, results from this study shed new light on the capacity of information routinely available in observational data to identify lung cancer patients with ‘good’ vs. ‘poor’ PS. This is especially useful for researchers interested in leveraging existing observational databases for comparative effectiveness research. Recent studies have highlighted likely overuse of chemotherapy in lung cancer treatment as well as aggressive treatment near end of life.²⁸⁻³⁰ Using a predictive model such as the one developed here with a threshold of 0.70 to proxy a patient as having ‘poor’ PS would ensure reasonably high

specificity (0.69 if PS = 2 considered ‘poor’) and thereby enable identification of a population for whom the receipt of chemotherapy appears inadvisable or requiring a more tailored discussion of less benefit and more risk per current guideline recommendations, and for whom early hospice intervention maybe warranted. On the other hand, using a lower predictive threshold (0.50) and thereby increasing the sensitivity of the predictive model may be useful to health disparities researchers, where interest might be in testing a hypothesis centered on under-treatment among minority populations. Similarly, choosing a predictive threshold with a high sensitivity could facilitate population identification for observational comparative effectiveness research. The best selection of both a predictive threshold and the allocation of PS = 2 patients will ultimately depend on the user’s objectives.

Performance status has long been considered one of the strongest prognostic factors³¹ and is used today by clinicians to assess the appropriateness of chemotherapy treatment and regimen choice for lung cancer patients.²² With the aging population, the number of Americans with functional limitations will increase dramatically, and therefore the urgency to capture and classify functional status information will grow.³² Furthermore, given the current challenges faced by the US health care system to deliver better and more cost-effective outcomes, the importance of comparative effectiveness studies is likely to only grow. Results from our study for the first time provide health services researchers and others with a viable tool to predict PS among lung cancer patients using information routinely available in observational data. As such, the value of observational data for comparative effectiveness research and for use by those interested in understanding cancer care quality or targeting specific lung cancer patients for possible inclusion in clinical trials, hospice care or other interventions is greatly enhanced.

Acknowledgments

This work was supported in part by grants from the Fund for Henry Ford and the National Cancer Institute. We thank Elizabeth Dobie and Nonna Akkerman for their assistance with data acquisition.

This research was supported by the Fund for Henry Ford and the National Cancer Institute, National Institutes of Health, U.S. under grant (NIH R01 CA114204-03).

References

1. Clinical practice guidelines for the treatment of unresectable non-small-cell lung cancer. Adopted on May 16, 1997 by the American Society of Clinical Oncology. J Clin Oncol. 1997 August; 15(8): 2996–3018. [PubMed: 9256144]
2. National Cancer Institute: Physician Data Query Cancer Information Summaries. National Cancer Institute. 2010 March 10 [Accessed May 11, 2010]. Available at: URL: <http://www.nci.nih.gov/cancerinfo/pdq/treatment/non-small-cell-lung/healthprofessional/>
3. Pfister DG, Johnson DH, Azzoli CG, et al. American Society of Clinical Oncology treatment of unresectable non-small-cell lung cancer guideline: update 2003. J Clin Oncol. 2004 January 15; 22(2):330–353. [PubMed: 14691125]
4. National Comprehensive Cancer Network and American Cancer Society: Lung Cancer: Treatment Guidelines for Patients. Version 1, December 2001. National Comprehensive Cancer Network. 2010 March 5. Available at: URL: <http://www.nccn.org>.
5. Smith TJ, Penberthy L, Desch CE, et al. Differences in initial treatment patterns and outcomes of lung cancer in the elderly. Lung Cancer. 1995 December; 13(3):235–252. [PubMed: 8719064]
6. Bach PB, Cramer LD, Warren JL, Begg CB. Racial differences in the treatment of early-stage lung cancer. N Engl J Med. 1999; 341(16):1198–1205. [PubMed: 10519898]
7. Earle CC, Venditti LN, Neumann PJ, et al. Who gets chemotherapy for metastatic lung cancer? Chest. 2000 May; 117(5):1239–1246. [PubMed: 10807806]

8. Potosky AL, Saxman S, Wallace RB, Lynch CF. Population variations in the initial treatment of non-small-cell lung cancer. *J Clin Oncol*. 2004 August 15; 22(16):3261–3268. [PubMed: 15310770]
9. Hillner BE, McDonald MK, Desch CE, Smith TJ, Penberthy LT, Retchin SM. A comparison of patterns of care of nonsmall cell lung carcinoma patients in a younger and Medigap commercially insured cohort. *Cancer*. 1998 November 1; 83(9):1930–1937. [PubMed: 9806651]
10. Bradley CJ, Dahman B, Given CW. Treatment and survival differences in older Medicare patients with lung cancer as compared with those who are dually eligible for Medicare and Medicaid. *J Clin Oncol*. 2008 November 1; 26(31):5067–5073. [PubMed: 18794546]
11. Scott WJ, Howington J, Feigenberg S, Movsas B, Pisters K. Treatment of non-small cell lung cancer stage I and stage II: ACCP evidence-based clinical practice guidelines (2nd edition). *Chest*. 2007 September; 132(3 Suppl):234S–242S. [PubMed: 17873171]
12. Blagden SP, Charman SC, Sharples LD, Magee LR, Gilligan D. Performance status score: do patients and their oncologists agree? *Br J Cancer*. 2003 September 15; 89(6):1022–1027. [PubMed: 12966419]
13. Buccheri G, Ferrigno D, Tamburini M. Karnofsky and ECOG performance status scoring in lung cancer: a prospective, longitudinal study of 536 patients from a single institution. *Eur J Cancer*. 1996 June; 32A(7):1135–1141. [PubMed: 8758243]
14. Lilenbaum RC, Cashy J, Hensing TA, Young S, Cella D. Prevalence of poor performance status in lung cancer patients: implications for research. *J Thorac Oncol*. 2008 February; 3(2):125–129. [PubMed: 18303431]
15. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *J Chron Dis*. 1987; 40(5):373–383. [PubMed: 3558716]
16. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *Journal of Clinical Epidemiology*. 1992 June; 45(6):613–619. [PubMed: 1607900]
17. Schneeweiss S, Seeger JD, Maclure M, Wang PS, Avorn J, Glynn RJ. Performance of comorbidity scores to control for confounding in epidemiologic studies using claims data. *Am J Epidemiol*. 2001 November 1; 154(9):854–864. [PubMed: 11682368]
18. AHFS Drug Information. Bethesda, MD: American Society of Health-System Pharmacists; 2010.
19. Efron B, Gong G. A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician*. February; 37(1):36–48. 83 A.D.
20. Lemeshow S, Hosmer DW Jr. A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol*. 1982 January; 115(1):92–106. [PubMed: 7055134]
21. Rodriguez E, Lilenbaum RC. New treatment strategies in patients with advanced non-small-cell lung cancer and performance status 2. *Clin Lung Cancer*. 2008 November; 9(6):326–330. [PubMed: 19073514]
22. Azzoli CG, Baker S Jr, Temin S, et al. American Society of Clinical Oncology Clinical Practice Guideline update on chemotherapy for stage IV non-small-cell lung cancer. *J Clin Oncol*. 2009 December 20; 27(36):6251–6266. [PubMed: 19917871]
23. Wright AA, Zhang B, Ray A, et al. Associations between end-of-life discussions, patient mental health, medical care near death, and caregiver bereavement adjustment. *JAMA*. 2008 October 8; 300(14):1665–1673. [PubMed: 18840840]
24. Zhang B, Wright AA, Huskamp HA, et al. Health care costs in the last week of life: associations with end-of-life conversations. *Arch Intern Med*. 2009 March 9; 169(5):480–488. [PubMed: 19273778]
25. Kelly K. Challenges in defining and identifying patients with non-small cell lung cancer and poor performance status. *Semin Oncol*. 2004 December; 31(6 Suppl 11):3–7. [PubMed: 15599828]
26. Zimmermann C, Burman D, Bandukwala S, et al. Nurse and physician inter-rater agreement of three performance status measures in palliative care outpatients. *Support Care Cancer*. 2010 May; 18(5):609–616. [PubMed: 19629537]
27. Iezzoni, LI. Risk adjustment for studying health care outcomes of people with disabilities. In: Iezzoni, LI., editor. *Risk Adjustment for Measuring Health Care Outcomes*. 3 ed.. Ann Arbor: Health Administration Press; 2003. p. 363–382.

28. Earle CC, Neville BA, Landrum MB, Ayanian JZ, Block SD, Weeks JC. Trends in the aggressiveness of cancer care near the end of life. *J Clin Oncol*. 2004 January 15; 22(2):315–321. [PubMed: 14722041]
29. Matsuyama R, Reddy S, Smith TJ. Why do patients choose chemotherapy near the end of life? A review of the perspective of those facing death from cancer. *J Clin Oncol*. 2006 July 20; 24(21): 3490–3496. [PubMed: 16849766]
30. Harrington SE, Smith TJ. The role of chemotherapy at the end of life: "when is enough, enough?". *JAMA*. 2008 June 11; 299(22):2667–2678. [PubMed: 18544726]
31. Albain KS, Crowley JJ, LeBlanc M, Livingston RB. Survival determinants in extensive-stage non-small-cell lung cancer: the Southwest Oncology Group experience. *J Clin Oncol*. 1991 September; 9(9):1618–1626. [PubMed: 1651993]
32. Iezzoni LI, Greenberg MS. Capturing and classifying functional status information in administrative databases. *Health Care Financ Rev*. 2003; 24(3):61–76. [PubMed: 12894635]

TABLE 1

The Eastern Cooperative Oncology Group (ECOG) Performance Status Score (PS) and its Karnofsky PS Equivalent

ECOG ¹ Grade	Description	Karnofsky ²
0	Fully active, able to carry on all pre-disease performance without restriction	100
1	Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature (e.g., light house work, office work)	80–90
2	Ambulatory and capable of all self care but unable to carry out any work activities. Up and about more than 50% of waking hours	60–70
3	Capable of only limited self care, confined to bed or chair more than 50% of waking hours	40–50
4	Completely disabled. Cannot carry on any self care; confined to bed or chair	20–30
5	Dead	0–10

(1) Oken MM, Creech RH, Tormey DC et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. *Am J Clin Oncol* 1982 December;5(6):649–55.

(2) Karnofsky DA, Burchenal JH. The Clinical Evaluation of Chemotherapeutic Agents in Cancer. In: MacLeod CM, editor. *Evaluation of Chemotherapeutic Agents*. Columbia Univ Press; 1949. p. 196.

TABLE 2

Sample Characteristics, and Prescription Drug and Medical Care Utilization in the Year Prior to Diagnosis of Lung Cancer, by Performance Status (PS) documentation level, (n = 552)

Characteristic	Unknown PS (n = 110)	Known PS (n = 442)	Documented PS (n = 261)	Extrapolated PS (n = 181)
Demographic Characteristics				
Average diagnosis age (SD)	66.9 (9.9)	67.5 (8.8)	68.0 (8.6)	66.9 (9.2)
Gender (%)				
Female	45	42	38*	48*
Male	55	58	62	52
Race (%)				
Asian	0	1	1	0
Black	31	30	28	34
White	69	69	71	66
Clinical Characteristics				
AJCC stage (%)				
II	4	11	10	11
IIIA	17	20	19	22
IIIB	24	18	17	20
IV	54	51	54	47
Average Charlson score (SD)	2.7 (3.6)	2.8 (3.3)	2.9 (3.5)	2.7 (3.1)
Atherosclerotic cardiovascular disease (%)	15.4	21.3	24.5*	16.6*
Congestive heart failure (%)	16.4	17.4	18.4	16.0
Ischemic heart disease (%)	8.2	9.3	10.3	7.7
Peripheral vascular disease (%)	11.8	13.4	14.2	12.2
Dementia (%)	0.9	1.1	1.5	0.6
Pelvic ulcer disease (%)	2.7	3.6	3.8	3.3
Rheumatologic disease (%)	5.4	6.3	6.1	6.6
Chronic pulmonary disease (%)	32.7	41.4	41.4	41.4
Liver disease (%)	2.7	1.4	0.8	2.2
Diabetes (%)	19.1	28.3	29.9	26.0
Diabetes with complications (%)	4.6	4.5	4.2	5.0
Paralysis (%)	1.8	0.9	1.1	0.6
AIDS (%)	1.8	1.1	1.1	1.1
Cancer (%)	24.6	29.9	28.0	32.6
Cancer with metastasis (%)	5.4	7.5	8.0	6.6
Renal disease (%)	6.4	5.7	6.1	5.0
Aneurysm (%)	6.4	6.3	7.3	5.0
Gangrene (%)	0.9	0.7	0.8	0.6
Prescription Drug Utilization				
Average no. of dispensings (SD)	8.7 (7.0)	9.4 (7.1)	9.5 (7.1)	9.3 (7.2)
Pct with ≥ 1 dispensing	85	89	91	86

Characteristic	Unknown PS (n =110)	Known PS (n = 442)	Documented PS (n = 261)	Extrapolated PS (n = 181)
Medical Care Utilization				
Average no. IP days (SD) ^a	2.7 (6.8)	3.0 (7.7)	2.4 (5.1) *	3.9(10.4) *
Average no. IP stays (SD)	0.4 (0.7)	0.5 (0.8)	0.4 (0.7) *	0.6 (0.8) *
Average length of IP stays (SD) ^b	5.8 (5.2)	6.1 (5.2)	6.0 (5.3)	6.1 (5.0)
Average no. of OP visits (SD)	5.9 (8.3)	5.6 (8.6)	5.7 (9.1)	5.5 (7.8)
Pct with ≥ 1 OP visit	75	76	77	76
Average ≥ 1 ED visit (SD)	0.7 (1.0)	0.6 (1.1)	0.7 (1.2)	0.5 (1.0)
Pct with ≥ 1 home health claim	25.4	29.0	31.4	25.4
Pct with ≥ 1 ambulatory surgery claim	2.7	3.2	3.4	2.8
Pct with ≥ 1 rehabilitation/SNF claim	4.6	3.8	5.0	2.2
Pct with ≥ 1 DME claim ^c	10.9	12.0	13.4	9.9

^a Among all patients (including those with no inpatient stays).

^b No. of inpatient days divided by no. of inpatient stays – among patients with at least one inpatient stay.

^c DME included claims for portable oxygen, walkers, canes, wheelchairs, and hospital beds.

* Significant difference at 5% level.

Abbreviations: standard deviation (SD); percent (Pct); number (no.); American Joint Committee on Cancer (AJCC); inpatient (IP); outpatient (OP); emergency department (ED); skilled nursing facility (SNF); durable medical equipment (DME).

TABLE 3

Among Patients with Known Performance Status (PS), Sample Characteristics, and Prescription Drug and Medical Care Utilization in the Year Prior to Diagnosis of Lung Cancer, by PS (n=442)

Characteristic	Pre 2009 ^a		Post 2009 ^b	
	Good PS (n = 290)	Poor PS (n = 152)	Good PS (n = 336)	Poor PS (n = 106)
Demographic Characteristics				
Average diagnosis age (SD)	66.4 (9.1) *	69.7 (7.9) *	66.8 (8.9) *	70.0 (8.1) *
Gender (%)				
Female	46 *	34 *	43	38
Male	54	66	57	62
Race (%)				
Asian	1	1	1	0
Black	28	34	28	38
White	71	65	71	62
Clinical Characteristics				
AJCC Stage (%)				
II	13	7	11	9
IIIA	24	13	23	11
IIIB	19	16	19	17
IV	44 *	64 *	47 *	63 *
Average Charlson score (SD)	2.4 (3.0) *	3.6 (3.8) *	2.5 (3.1) *	3.9 (3.9) *
Atherosclerotic cardiovascular disease (%)	19.3	25.0	19.9	25.5
Congestive heart failure (%)	12.1 *	27.6 *	14.3 *	27.4 *
Ischemic heart disease (%)	6.6 *	14.5 *	7.7 *	14.2 *
Peripheral vascular disease (%)	10.3 *	19.1 *	12.2	17.0
Dementia (%)	0.0 *	3.3 *	0.0 *	4.7 *
Pelvic ulcer disease (%)	2.1 *	6.6 *	2.4 *	7.6 *
Rheumatologic disease (%)	5.9	7.2	6.3	6.6
Chronic pulmonary disease (%)	34.5 *	54.6 *	35.4 *	60.4 *
Liver disease (%)	1.0	2.0	1.0	3.0
Cancer (%)	28.3	32.9	27.7	36.8
Cancer with metastasis (%)	6.6	9.2	6.8	9.4
Diabetes (%)	26.2	32.2	28.0	29.2
Diabetes with complications (%)	3.8	5.9	3.9	6.6
Paralysis (%)	0.3	2.0	0.3	2.8
AIDS (%)	1.0	1.3	0.9	1.9
Renal disease (%)	3.8 *	9.2 *	3.9 *	11.3 *
Aneurysm (%)	5.2	8.6	6.2	6.6
Gangrene (%)	0.3	1.3	0.3	1.9
Prescription Drug Utilization				

Characteristic	Pre 2009 ^a		Post 2009 ^b	
	Good PS (n = 290)	Poor PS (n = 152)	Good PS (n = 336)	Poor PS (n = 106)
Average no. of dispensings (SD)	9.0 (7.1)	10.3 (7.2)	9.1 (7.1)	10.5 (7.3)
Pct with ≥ 1 dispensing	87	93	88	92
Medical Care Utilization				
Average no. IP days (SD) ^c	1.7 (3.9) *	5.5 (11.7) *	1.8 (4.1) *	6.8 (13.3) *
Average no. IP stays (SD) ^d	0.3 (0.6)	0.8 (1.0)	0.3 (0.6)	0.9 (1.1)
Average length of IP stays (SD)	5.4 (4.8)	6.8 (5.5)	5.7 (5.1)	6.7 (5.2)
Average no. of OP visits (SD)	6.0 (8.5)	4.9 (8.7)	6.0 (8.6)	4.5 (8.6)
Pct with ≥ 1 OP visit	82 *	66 *	81 *	63 *
Average ≥ 1 ED visit (SD)	0.5 (0.9) *	0.8 (1.3) *	0.5 (1.0) *	0.9 (1.3) *
Pct with ≥ 1 home health claim	23.1 *	40.1 *	24.7 *	42.4 *
Pct with ≥ 1 ambulatory surgery claim	3.8	2.0	3.6	1.9
Pct with ≥ 1 rehabilitation/SNF claim	3.1	5.3	3.3	5.7
Pct with ≥ 1 DME claim ^e	6.9 *	21.7 *	8.0 *	24.5 *

^aPre 2009: Good PS = ECOG: 0 – 1; KPS: 80 – 100; Poor PS = ECOG: 2 – 5; KPS: 0 – 70.

^bPost 2009: Good PS = ECOG: 0 – 2; KPS: 60 – 100; Poor PS = ECOG: 3 – 5; KPS: 0 – 50.

^cAmong all patients (including those with no inpatient stays).

^dNo. of inpatient days divided by no. of inpatient stays – among patients with at least one inpatient stay.

^eDME included claims for portable oxygen, walkers, canes, wheelchairs, and hospital beds.

* Significant difference at 5% level.

Abbreviations: standard deviation (SD); percent (Pct); number (no.); American Joint Committee on Cancer (AJCC); inpatient (IP); outpatient (OP); emergency department (ED); skilled nursing facility (SNF); durable medical equipment (DME).

Estimated Logistic Regression Parameters (β) and Standard Errors (SE) and Measures of Model Performance for Alternative Models of Performance Status Predictors

TABLE 4

PS (Pre 2009): Good/Poor ^a	Model 1		Model 2		Model 3	
Variable	β	SE	β	SE	β	SE
Intercept	3.97	0.85***	3.65	0.94***	4.55	1.03***
Age in years at diagnosis	-0.04	0.01**	-0.04	0.01**	-0.04	0.01**
AJCC Stage IV	-0.84	0.21***	-0.81	0.23**	-0.84	0.23**
Chronic pulmonary disease			-0.67	0.24**	-0.63	0.24**
Number of Inpatient Stays			-0.58	0.16**	-0.60	0.16**
Any number of outpatient visits			1.05	0.26***	1.19	0.27***
Number of ED visits			-0.22	0.10*	-0.22	0.10*
Any number of DME claim			-0.80	0.35*	-0.81	0.36*
Any prescription drug dispensing					-1.08	0.43*
Model Performance						
C statistic (95% CI)	0.66 (0.61, 0.71)		0.75 (0.71, 0.81)		0.76 (0.72, 0.81)	
R ² (95% CI)	0.07 (0.03, 0.11)		0.19 (0.13, 0.26)		0.20 (0.15, 0.28)	
Predictive Threshold = 0.50						
Sensitivity (95% CI)	0.88 (0.78, 1.00)		0.90 (0.84, 0.94)		0.88 (0.84, 0.93)	
Specificity (95% CI)	0.20 (0.00, 0.40)		0.43 (0.31, 0.55)		0.45 (0.34, 0.56)	
False positive (95% CI)	0.32 (0.28, 0.37)		0.25 (0.21, 0.29)		0.25 (0.20, 0.29)	
False negative (95% CI)	0.54 (0.29, 0.87)		0.31 (0.24, 0.40)		0.34 (0.25, 0.40)	
Predictive Threshold = 0.60						
Sensitivity (95% CI)	0.72 (0.61, 0.83)		0.79 (0.72, 0.87)		0.79 (0.72, 0.87)	
Specificity (95% CI)	0.43 (0.27, 0.65)		0.53 (0.45, 0.66)		0.55 (0.47, 0.67)	
False positive (95% CI)	0.29 (0.23, 0.33)		0.24 (0.18, 0.27)		0.23 (0.18, 0.26)	
False negative (95% CI)	0.55 (0.46, 0.62)		0.43 (0.34, 0.49)		0.42 (0.34, 0.48)	
Predictive Threshold = 0.70						
Sensitivity (95% CI)	0.49 (0.33, 0.65)		0.61 (0.52, 0.74)		0.64 (0.53, 0.75)	
Specificity (95% CI)	0.75 (0.60, 0.88)		0.69 (0.60, 0.79)		0.69 (0.61, 0.79)	

PS (Pre 2009): Good/Poor ^a						
Variable	Model 1		Model 2		Model 3	
	β	SE	β	SE	β	SE
False positive (95% CI)	0.21	(0.16, 0.26)	0.21	(0.16, 0.25)	0.20	(0.15, 0.25)
False negative (95% CI)	0.56	(0.51, 0.62)	0.52	(0.44, 0.57)	0.50	(0.43, 0.56)
PS (Pre 2009): Good/Poor ^b						
Variable	Model 1		Model 2		Model 3	
	β	SE	β	SE	β	SE
Intercept	4.39	0.94***	4.25	1.08***	5.08	1.18***
Age in years at diagnosis	-0.04	0.01***	-0.04	0.02**	-0.04	0.02**
AJCC Stage IV	-0.64	0.23***	-0.55	0.26**	-0.57	0.26**
Chronic pulmonary disease			-0.88	0.27***	-0.83	0.27***
Number of Inpatient Stays			-0.80	0.17***	-0.81	0.17***
Any number of outpatient visits			1.08	0.29***	1.20	0.30***
Number of ED visits			-0.25	0.11**	-0.25	0.11**
Any number of DME claim			-0.60	0.36*	-0.61	0.36*
Any prescription drug dispensing					-0.98	0.50*
Model Performance						
C statistic (95% CI)	0.64	(0.58, 0.70)	0.78	(0.74, 0.84)	0.78	(0.74, 0.85)
R ² (95% CI)	0.04	(0.01, 0.08)	0.19	(0.13, 0.28)	0.20	(0.14, 0.28)
Predictive Threshold = 0.50						
Sensitivity (95% CI)	1.00	(0.98, 1.00)	0.94	(0.91, 0.97)	0.94	(0.91, 0.97)
Specificity (95% CI)	0.00	(0.00, 0.01)	0.28	(0.19, 0.47)	0.32	(0.21, 0.50)
False positive (95% CI)	0.24	(0.20, 0.28)	0.19	(0.15, 0.22)	0.19	(0.14, 0.22)
False negative (95% CI)	1.00	(0.98, 1.00)	0.40	(0.23, 0.48)	0.38	(0.21, 0.46)
Predictive Threshold = 0.60						
Sensitivity (95% CI)	0.96	(0.87, 1.00)	0.90	(0.86, 0.94)	0.90	(0.86, 0.94)
Specificity (95% CI)	0.06	(0.00, 0.27)	0.42	(0.29, 0.59)	0.45	(0.32, 0.60)
False positive (95% CI)	0.24	(0.20, 0.27)	0.17	(0.13, 0.20)	0.16	(0.12, 0.20)
False negative (95% CI)	0.71	(0.37, 1.00)	0.42	(0.32, 0.52)	0.42	(0.32, 0.51)
Predictive Threshold = 0.70						

PS (Pre 2009): Good/Poor ^a		Model 1		Model 2		Model 3	
Variable		β	SE	β	SE	β	SE
Sensitivity (95% CI)		0.76 (0.62, 0.91)		0.83 (0.78, 0.99)		0.83 (0.76, 0.88)	
Specificity (95% CI)		0.37 (0.13, 0.63)		0.56 (0.22, 0.72)		0.55 (0.46, 0.70)	
False positive (95% CI)		0.21 (0.16, 0.24)		0.14 (0.11, 0.21)		0.15 (0.10, 0.17)	
False negative (95% CI)		0.67 (0.55, 0.78)		0.49 (0.28, 0.58)		0.50 (0.41, 0.57)	

Model 1: Significant ($P<0.05$) performance status predictors from tumor registries (age at diagnosis and AJCC stage). Model 2: Significant performance status predictors from medical claims (age at diagnosis, diagnosis of COPD, inpatient stays, any outpatient visits, and emergency department visit). Model 3: Significant performance status predictors from tumor registries, medical claims, and pharmacy claims (age at diagnosis, AJCC stage, diagnosis of COPD, inpatient stays, any outpatient visits, emergency department visit, and any prescriptions).

* $P<0.05$;
** $P<0.01$;
*** $P<0.0001$.

Abbreviations: American Joint Committee on Cancer (AJCC); emergency department (ED); durable medical equipment (DME); chronic pulmonary disease (COPD).

^aPre 2009: Good PS = ECOG: 0 – 1 or KPS: 80 – 100 / Poor PS = ECOG: 2 – 5 or KPS: 0 – 70.

^bPost 2009: Good PS = ECOG: 0 – 2 or KPS: 60 – 100 / Poor PS = ECOG: 3 – 5 or KPS: 0 – 50.

TABLE 5

Number of Patients with Observed (Predicted) Good Performance Status (PS) by Model and Model Determined Decile^a

Rank Deciles	Number with Observed (Predicted) Good PS (Pre 2009) ^b		
	Model 1	Model 2	Model 3
1	23 (21)	10 (9)	8 (8)
2	23 (22)	13 (18)	14 (18)
3	25 (24)	29 (24)	28 (23)
4	26 (27)	29 (27)	27 (27)
5	23 (27)	31 (30)	32 (30)
6	28 (31)	33 (32)	37 (33)
7	26 (29)	34 (34)	31 (36)
8	37 (34)	33 (36)	34 (37)
9	39 (37)	37 (40)	38 (39)
10	40 (38)	41 (40)	41 (39)
<i>P</i> value	0.69	0.32	0.13

Rank Deciles	Number with Observed (Predicted) Good PS (Post 2009) ^c		
	Model 1	Model 2	Model 3
1	28 (28)	12 (12)	10 (12)
2	29 (30)	22 (24)	24 (24)
3	34 (33)	32 (30)	32 (30)
4	37 (36)	33 (33)	34 (33)
5	32 (34)	38 (36)	36 (36)
6	34 (35)	40 (38)	37 (37)
7	39 (37)	35 (38)	39 (39)
8	42 (42)	42 (40)	39 (40)
9	36 (37)	41 (42)	41 (42)
10	25 (23)	41 (41)	44 (43)
<i>P</i> value	0.92	0.63	0.98

^aDeciles were created by ranking patients according to increasing predicted likelihood of good performance status on the basis of the explanatory variables in each of the three models.

^bPre 2009: Good PS = ECOG: 0 – 1 or KPS: 80 – 100 / Poor PS = ECOG: 2 – 5 or KPS: 0 – 70.

^cPost 2009: Good PS = ECOG: 0 – 2 or KPS: 60 – 100 / Poor PS = ECOG: 3 – 5 or KPS: 0 – 50.