Evaluating claims-based indicators of the intensity of end-of-life cancer care

CRAIG C. EARLE¹, BRIDGET A. NEVILLE¹, MARY BETH LANDRUM², JEFFREY M. SOUZA², JANE C. WEEKS¹, SUSAN D. BLOCK³, EVA GRUNFELD⁴ AND JOHN Z. AYANIAN²,⁵

¹Division of Population Sciences, Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA, ²Department of Health Care Policy, Harvard Medical School, Boston, MA, USA, ³Division of Psychosocial Oncology and Palliative Care, Department of Psychiatry, Brigham and Women’s Hospital, Boston, MA, USA, ⁴Dalhousie University and Cancer Care Nova Scotia, Halifax, Nova Scotia, Canada, and ⁵Division of General Medicine, Department of Medicine, Brigham and Women’s Hospital, Boston, MA, USA

Abstract

Objective. To evaluate measures that could use existing administrative data to assess the intensity of end-of-life cancer care.

Methods. Benchmarking standards and statistical variation were evaluated using Medicare claims of 48,906 patients who died from cancer from 1991 through 1996 in 11 regions of the United States. We assessed accuracy by comparing administrative data to 150 medical records in one hospital and affiliated cancer treatment center.

Results. Systems not providing overly aggressive care near the end of life would be ones in which less than 10% of patients receive chemotherapy in the last 14 days of life, less than 2% start a new chemotherapy regimen in the last 30 days of life, less than 4% have multiple hospitalizations or emergency room visits or are admitted to the intensive care unit (ICU) in the last month of life, and less than 17% die in an acute care institution. At least 55% of patients would receive hospice services before death from cancer, and less than 8% of those would be admitted to hospice within only 3 days of death. All measures were found to have accuracy ranging from 85 to 97% and 2- to 5-fold adjusted variability between the 5th and 95th percentiles of performance.

Conclusions. The usefulness of these measures will depend on whether the concept of intensity of care near death can be further validated as an acceptable and important quality issue among patients, their families, health care providers, and other stakeholders in oncology.

Keywords: administrative data, benchmarking, cancer, claims data, indicators development, performance measures, quality of care, terminally ill patients

Most of the quality of care measurement work conducted to date in oncology has focused either on screening or the initial treatment of early-stage disease, and studies of breast cancer dominate this literature [1]. Unfortunately, however, almost half of cancer patients either present with, or eventually develop, incurable metastatic disease [2]. Despite this pattern, there is a ‘dramatic lack of data’ available about the quality of cancer care for patients with advanced disease [3]. To address this concern, we have previously reported a qualitative study which identified a set of promising performance measures that use administrative data sources: processes of overuse of chemotherapy, underuse of hospice services, and the outcomes of frequent emergency room visits, hospitalizations, and intensive care unit (ICU) admissions (possibly indicating misuse of aggressive intervention) near the end of life [4]. Moreover, an evaluation of these performance measures in Medicare claims for 8,155 patients over age 65 with terminal lung, breast, colorectal, and other gastrointestinal cancers found steadily increasing use of chemotherapy very near death over a 4-year span between 1993 and 1996, a trend that was also associated with a rise in potentially negative processes of care such as increased emergency room visits and ICU admissions and delayed hospice admission [5]. These processes were more likely to be experienced by African-American patients and patients living in areas with less hospice services.

In this study, we assess the statistical properties of these measures to determine empirical benchmarking standards and to evaluate whether they have sufficient accuracy and variation to merit further development as performance measures for end-of-life cancer care.

Methods

As previously reported, we conducted a comprehensive literature review and focus groups with patients and bereaved family members to identify potential performance measures. An expert panel of health care providers subsequently used a modified Delphi approach to rank the measures based on meaningfulness and importance [4]. Candidate measures were...
then assessed for feasibility: the numerator and denominator were defined for each measure, and analytic procedures were created to operationalize them in administrative data, specifically in Medicare claims and local billing data.

Here we report three further methodologic evaluations [6]:

1. Establishing achievable benchmarks of care.
2. Determining the accuracy of the performance measures.
3. Assessing practice variability.

**Establishing performance measures and achievable benchmarks of care**

**Data sources.** During the period between 1991 and 1996, the National Cancer Institute’s (NCI) Surveillance, Epidemiology, and End Results (SEER) program included eleven participating tumor registries: San Francisco/Oakland, Connecticut, Detroit, Hawaii, Iowa, New Mexico, Seattle/Puget Sound, Utah, Atlanta, San Jose/Monterey, and Los Angeles. These registries collect uniform information on all cancers diagnosed within their geographic regions, capturing about 97% of all incident cases in those areas [7]. The geographic areas covered by SEER contain approximately 14% of the American population [8] and are demographically fairly representative [9]. Patients eligible for Medicare represent about half of all cases in these regions. Disease-related data collected include the cancer site, stage, histology, date of diagnosis, and date and cause of patient death. The registries also record patient socio-demographic characteristics such as age, sex, and race and link census data on each patient’s census tract or zip code of residence, including median and per capita income and wealth, educational levels, and racial mix. The Centers for Medicare and Medicaid Services’ (CMS, formerly the Health Care Financing Administration, HCFA) Medicare database includes files for in-patient and outpatient care, physician and laboratory billings, as well as bills for home health and hospice care. Each patient in the SEER and Medicare databases has a unique case identification number that permits matching and merging of the different files. In this way, cases over age 65 have been linked between the databases with a 94% match rate [10].

**Cohort identification.** The study sample consisted of 48,906 decedents representing all Medicare eligible patients over age 65 who were identified on the death certificate as having died from lung, breast, colorectal, or other gastrointestinal cancers, diagnosed while residing in one of the 11 SEER areas between January 1, 1991 and December 31, 1996, with 1996 being the most recent year with complete survival data in SEER at the time of analysis. Patients were excluded if they were not enrolled in both Part A (5.1%) and Part B (6.1%) of Medicare at any time during the study period, as complete treatment information is not available for these patients. Patients were also excluded if their Medicare entitlement was on the basis of disability or end-stage renal disease (5.2%), if the dates of diagnosis or death differed by more than 2 months in the SEER and Medicare databases (0.04%), or if their cancer was first identified at the time of death or autopsy (0.26%). As these are not mutually exclusive categories, overall 11.5% of patients were excluded for these reasons. Also removed were those who were ever enrolled in an Health Maintenance Organisation (HMO) at any time during the study period (15.3%).

**Definitions of explanatory variables.** Patients were classified by age at diagnosis, race (White, Black, or other), and ethnicity (Hispanic versus non-Hispanic). Socioeconomic quintiles were developed based on the availability of information, according to the following hierarchy of decreasing specificity: (i) race- and age-specific median household income by census tract; (ii) unadjusted median household income by census tract; (iii) median household wealth by census tract; and (iv) median household wealth by zip code.

Use of chemotherapy was identified from billing claims using standard algorithms that combine ICD-9-CM (International Classification of Disease and Clinical Modification) codes with Healthcare Common Procedure Coding System (HCPCS) codes, Current Procedural Terminology (CPT) codes, Diagnosis Related Group (DRG) codes, and revenue center codes [11]. We also used those codes to capture additional radiation treatments from billing claims. We calculated the Charlson comorbidity index [12,13] by identifying whether certain ICD-9 codes were recorded in months −2 to −14 before the diagnosis of cancer in each patient, following the method described by Deyo [14]. We used both in-patient and outpatient bills to derive comorbidity scores, excluding cancer-related codes and required a diagnosis to appear on at least two occasions separated by at least a month, as suggested by Klabunde [15]. We analyzed the Charlson score as an ordered-categorical variable (scores grouped 0, 1, 2, and >2) [12]. Care in a teaching hospital was identified if there was a bill for Indirect Medical Education at any time during the patient’s disease course [14,15].

**Geographic unit of analysis: health care service area of residence.** Health care service areas (HCSAs) are groupings of metropolitan statistical areas defined by observed patient flow patterns in Medicare [16]. To assess characteristics of the 77 HCSAs comprising the SEER areas, we used data from the National Center for Health Workforce Information and Analysis’ Area Resource File to calculate the following HCSA characteristics: the per capita density of physicians, medical specialists, and radiation oncologists derived from theAMA Master File; and the density of hospitals, teaching hospitals, hospital beds, hospitals with oncology services, and hospices from the County Hospital File.

**Establishing achievable benchmarks of care.** We next created performance measures based on empirical benchmarking. This was done by measuring the performance of all providers and using the rate achieved by the top performers as an achievable standard. Because it is data driven, benchmarking has the advantage of setting targets that are demonstrably attainable [17]. Furthermore, by repeatedly measuring performance and changing the standard, continuous quality improvement should result.

We considered each HCSA to contain a related group of providers. We calculated benchmark rates with the ‘pared-mean method.’ Using this approach, we defined the ‘performance fraction’ as the number of patients experiencing the performance measure event, divided by the total number of patients eligible for the service. HCSAs were ranked according to
observed performance on each of the measures and then grouped into deciles of the population. The benchmark was then calculated for the decile corresponding to the best performing HCSAs as

\[
\text{Benchmark} = \frac{\text{Total number of patients in the 10\% subset experiencing the performance measure}}{\text{Total number of patients in the 10\% subset}}
\]

**Accuracy of the measures**

Assessment of the accuracy and reliability of the claims-based candidate performance measures involved determining whether administrative data and coding of the relevant processes were acceptably accurate and complete. The Dana-Farber/Partners Cancer Care Consolidated Tumor Registry captures all patients diagnosed with cancer who are treated at the Brigham and Women’s Hospital or Dana-Farber Cancer Institute. We obtained the medical records of 150 consecutive patients treated within this health care system who died of one of the four tumor types evaluated in the SEER-Medicare claims. Patients’ medical records were reviewed and the occurrence of each service and its dates were abstracted and recorded on a data abstraction form. These data were then entered into an Access database which was linked through patients’ hospital numbers to in-patient and outpatient claims from each institution. Billing claims were compared with medical records, with the medical records considered the gold standard. The administrative data were considered to be accurate if they correctly identified the occurrence of one of the processes being measured on a date that was within one calendar day of when the process was recorded in the medical record. From this, the sensitivity, specificity, and ‘% accurate’ were calculated.

**Practice variability**

If there is not enough variability in practice, then a performance measure will not be able to detect differences in the quality of care [18]. We assessed variability in the candidate performance measures by estimating the variability across HCSAs in the SEER-Medicare data set using hierarchical regression models. All models were fit using BUGS software [19] and were case-mix adjusted for age, sex, race, socioeconomic status, comorbidity, disease site and stage, care in a teaching hospital, urban/rural residence, and year of death. Variation was then expressed as the adjusted rate of the performance measure in the HCSA at the 5th percentile relative to the adjusted rate in the HCSA at the 95th percentile [20].

**Results**

**Benchmarks**

Table 1 summarizes the results of the benchmarking assessments. Assuming that the 10th decile represents a desirable benchmark, the analysis of SEER-Medicare claims suggests that health care systems not providing overly aggressive care would be ones in which less than 10% of patients receive chemotherapy in the last 14 days of life, less than 2% start a new chemotherapy regimen in the last 30 days of life, less than 4% have multiple hospitalizations or emergency room visits or are admitted to the ICU in the last month of life, and less than 17% die in an acute care institution. At least 55% of patients would receive hospice services before death from cancer, and less than 8% of those would be admitted to hospice within only 3 days of death.

**Accuracy**

Overall, we found the administrative data from one institution to be remarkably accurate in most cases, with all performance measures exceeding our prespecified cutoff of 0.75 (Table 1). Some problems were noted, however. In-patient chemotherapy might be missed if it was not the reason for the admission. For example, patients admitted for a complication of their disease or treatment might be given chemotherapy

<table>
<thead>
<tr>
<th>Performance measure</th>
<th>Benchmark</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Accuracy</th>
<th>Variability (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion receiving chemotherapy in the last 14 days of life</td>
<td>&lt;0.10</td>
<td>0.92</td>
<td>0.94</td>
<td>0.92</td>
<td>2.24 (1.74–2.97)</td>
</tr>
<tr>
<td>Proportion starting a new chemotherapy regimen in the last 30 days of life</td>
<td>&lt;0.02</td>
<td>0.83</td>
<td>0.94</td>
<td>0.85</td>
<td>3.19 (2.03–5.41)</td>
</tr>
<tr>
<td>&gt;1 emergency room visit in the last month of life</td>
<td>&lt;0.04</td>
<td>0.82</td>
<td>0.96</td>
<td>0.89</td>
<td>2.78 (2.04–3.88)</td>
</tr>
<tr>
<td>&gt;1 hospitalization in the last month of life</td>
<td>&lt;0.04</td>
<td>0.96</td>
<td>1.00</td>
<td>0.97</td>
<td>2.38 (1.85–3.16)</td>
</tr>
<tr>
<td>Admission to the ICU in the last month of life</td>
<td>&lt;0.04</td>
<td>0.87</td>
<td>0.97</td>
<td>0.95</td>
<td>3.28 (2.38–4.67)</td>
</tr>
<tr>
<td>Death in an acute care hospital</td>
<td>&lt;0.17</td>
<td>0.95</td>
<td>1.00</td>
<td>0.97</td>
<td>2.49 (2.05–3.12)</td>
</tr>
<tr>
<td>Lack of admission to hospice</td>
<td>&lt;0.45</td>
<td>0.24</td>
<td>0.96</td>
<td>0.88</td>
<td>5.00 (3.76–6.89)</td>
</tr>
<tr>
<td>Admission to hospice &lt;3 days before death</td>
<td>&lt;0.08</td>
<td>0.97</td>
<td>1.00</td>
<td>0.97</td>
<td>2.39 (1.99–2.95)</td>
</tr>
</tbody>
</table>

Accuracy, the % agreement within +/- 1 day; benchmark, the performance of the top decile of health care service areas (HCSA); CI, confidence interval; ICU, intensive care unit; sensitivity and specificity refers to claims, compared with medical record review as the gold standard; variability, ratio of adjusted rates in the 5th and 95th percentile HCSAs.
before discharge if it was due, but that was not always reflected in in-patient codes. Because our local hospital is a teaching hospital, patients were occasionally seen and discharged from the emergency department without being seen by an attending physician. In those cases, appropriately, no bill was generated. ICU admission and discharge dates are not actually provided in standard claims, so the dates of hospitalization and ICU length of stay are used as a proxy, leading to a lack of accuracy. Hospice is captured poorly in both institutional claims and medical records. Bills in our local hospital system could effectively only identify when a patient was discharged from hospital to an in-patient hospice, and even then such a discharge could be confused with a rehabilitation program admission. Hospice is often arranged without a clinic visit, and consequently not captured in the medical record. Claims from insurers (Medicare or an HMO) may actually be more complete in this area, however. Except for hospice care, administrative data tended to underestimate the aggressiveness of care.

Variability

All performance measures were found to have statistically significant variability between health care service areas in hierarchical regression models (Table 1). The adjusted variation in performance measure rates between areas at the 5th and 95th percentiles ranged from 2.24- to 5-fold. Although the eight indicators were highly correlated with each other at the patient level, they were only moderately correlated at the area level. The largest correlations were between the two measures of intensive use of chemotherapy (correlation equal to 0.47, \( P < 0.001 \)) and between at least one hospitalization in the last month of life and death in the hospital (correlation equal to 0.43, \( P < 0.001 \)). To further examine structure among the measures at the area level, we performed an exploratory factor analysis. Examining factor loadings suggests that the eight measures partition roughly into one factor representing hospitalizations, ICU admissions and death in the hospital; a second factor representing the two chemotherapy measures; and a third factor representing ER visits, lack of hospice admission and short lengths of stay among hospice enrollees. These three factors could explain 63% of the variance among the eight measures. In contrast, a single factor was able to explain only 26% of the variance, suggesting at least three distinct dimensions of intensive treatment at the end of life.

Discussion

We have identified a series of candidate quality performance measures that can be applied to administrative data to profile cancer care near the end of life. In this study, we have taken an empirical approach to determine benchmarks for these measures and assess their variability and accuracy. Furthermore, these potential performance measures evaluate processes of care that are applicable to many types of cancer. Consequently, they can be assessed in large samples of patients in order to obtain stable estimates of performance. Their ultimate utility depends on whether the intensive use of aggressive treatment near death is perceived as suboptimal quality of care by patients, their families, health care providers, payers, the public, and policymakers.

Patterns of overly aggressive cancer care near the end of life may be a marker for situations in which providers are avoiding difficult discussions to prepare patients to accept terminal care. It is usually easier to simply recommend a new course of chemotherapy. Elements of prognostic skill or judgment resulting in injudicious use of treatments, with consequent high rates of complications, could also be at play. Alternatively, the performance measures may be reflecting a lack of available palliative and hospice resources, as we have previously reported [5]. This latter explanation is attractive as it leads to clear policy-related solutions. If validated, these measures could be used by health systems or insurers to identify areas where supportive services may be lacking, resulting in possible overuse of cancer-directed therapies near death.

There are some limitations to our analyses. The SEER-Medicare data set, which was used for much of this work, is restricted to patients over age 65. However, because cancer is commonly a disease of the elderly, over half of all cancer care in the United States is covered by Medicare. The SEER-Medicare database also represents only specific geographic locations and misses the 10–15% of patients enrolled in Medicare HMOs. The local database used for accuracy ascertainment was limited to only one urban academic hospital system and as such may not be representative of the billing practices in other settings. Most importantly, validation of these measures is needed to ensure that intensity of care is related to quality of care.

There are difficulties inherent in any effort to develop quality indicators for end-of-life care. There is a limited evidence base, little consensus among experts and patients as to what constitutes optimal care, and the end-of-life period is hard to identify prospectively. Looking backwards at dead patients is an imperfect proxy. In ongoing work, we are examining these performance measures in patients who die from other cancers, in younger patients, and in patients with other insurance types, including patients in a Canadian health care setting. We are also field testing the measures in private oncologists’ offices and validating them prospectively in large cohorts of patients with lung and colorectal cancer in the NCIf’s Cancer Care Outcomes Research and Surveillance Consortium (CanCORS) [21]. It is possible, however, that even with ideal data there may not be perfect correlation between prospective satisfaction and these performance measures because patients and families may not really understand when the patient is close to death. Moreover, they lack the perspective of having seen the disease through to its end. Attitudes may shift away from a preference for aggressive care only after the patient dies. As a result, it might not be possible to both maximize the satisfaction of patients approaching end-of-life, and avoid futile care. The challenge, then, is in establishing such measures for accountability or quality improvement will be to determine whether the concept of overuse of non-curative treatment near death can be accepted as an important quality issue among the various stakeholders in cancer care.
Acknowledgement

Supported by a grant from the National Cancer Institute (CA 91753–02).

References


Accepted for publication 26 May 2005