Ability of Medicare Claims Data and Cancer Registries to Identify Cancer Cases and Treatment

Donna Katzman McClish,1 Lynne Penberthy,1,2 Martha Whittemore,2 Craig Newschaffer,3 Diane Woolard,4 Christopher E. Desch,5,6 and Sheldon Retchin5

The objective of this study is to compare the ability of Medicare and cancer registry data to identify incident cancer cases and initial surgical therapy both singly and in combination. Data from the Virginia Cancer Registry (VCR) were linked to Medicare claims files (Medical Provider Analysis and Review File (MEDPAR)) for Virginia residents aged 65 years and over with breast, colorectal, lung, or prostate cancer diagnosed between 1986 and 1989. MEDPAR found 73–83% of cancer cases identified by VCR. Factors significantly associated with MEDPAR missing a case that was reported to VCR included younger age, male gender, living in an urban area, higher social class, in situ disease, and lack of cancer treatment. A total of 70–82% of cancer cases identified through Medicare claims were reported to the VCR. Older age, female gender, nonwhite race, comorbid conditions, no surgical procedures, multiple cancer admissions, and the position of the cancer diagnostic code on the MEDPAR record were factors significantly related to being missed by the VCR. The rate of capturing initial surgical therapies was similar to that of identifying cases. Combining information from VCR and MEDPAR resulted in increasing sensitivity for identifying incident cases to 92–97%. Using combined data from independent sources may improve reporting, increase the accuracy of cancer incidence estimates, and provide an opportunity to identify reasons for missing data. Am J Epidemiol 1997;145:227–33.

epidemiologic methods; incidence; Medicare; neoplasms; registries; sensitivity and specificity

There is great interest in evaluating incidence, costs, and patterns of cancer care in the United States. Data for studies of these topics frequently do not involve primary data collection, but rather administrative databases, which have some of the most readily available information on costs and resource utilization. Potential data sources for large-scale cancer studies include state and local central cancer registries, the Surveillance, Epidemiology, and End Results (SEER) program, Medicare and Medicaid data, as well as data from private insurers.

Administrative data have been used to study issues of reimbursement (1), disease incidence (2–6), and practice variation (7–9). More recently, administrative data, particularly Medicare claims data, have been used successfully for outcomes research (10–13). Questions have been raised about the usefulness and validity of administrative data for epidemiologic and other purposes (14–22). A number of studies have examined the accuracy of diagnostic coding in databases that contain hospital discharge data, including Medicare (8–22). The ability of Medicare data to represent the US elderly population has also been questioned (14). Even the representativeness of the SEER data has been challenged, although it covers 10 percent of the US population (5, 6, 22, 23). A few studies (5, 6, 22) have compared cancer incidence and surgical rates between Medicare and SEER data and have concluded that the incidence rates are often reasonably close. Resection rates were less accurate, often 12–27 percent lower. However, comparison of rates does not allow for characterization of the differences on the individual level. As stated by Whittle et al., "The accuracy of Part A and Part B data, alone or in combination, could best be assessed by linkages of SEER and Medicare data at the level of individual Medicare beneficiaries" (22, p. 1234).
The purpose of this paper is to compare in detail the ability of Medicare and cancer registry data to identify cancer cases and initial cancer treatment for the elderly in Virginia, both singly and in combination. Unlike prior studies, which have compared incidence rates at the population level, this study links data sources so that results can be assessed at the individual patient level. This enables the following questions to be addressed. How well do the Virginia Cancer Registry (VCR) and Medicare hospitalization data each identify incident cases of breast, colorectal, lung, and prostate cancer in people aged 65 years and over? Are the cases who were not identified missing at random, or does a systematic bias exist? How would using the two data sources together change estimates of cancer incidence? Finally, how well do the two sources identify initial surgical treatment for cancer?

MATERIALS AND METHODS

Data sources

This study is part of a larger one designed to examine costs and patterns of care for cancer in elderly Virginia residents. Sources of data include the VCR, the Medicare Annual Demographic Files, the Medicare Provider Analysis and Review (MEDPAR) files, and the 1990 Census Data for Zip Code Level Information (Tape 3B).

The VCR is a central cancer registry that has collected and maintained data on cancer cases since 1970. Between 1985 and 1989, the registry was not population based. Approximately 50 hospitals (nearly 85 percent of hospital beds) in Virginia reported on a voluntary basis for American College of Surgeons certification. For this reason, when we compared VCR and MEDPAR, only hospitals reporting to VCR were included. The VCR collects and maintains clinical data including site, summary stage, and initial therapy.

The MEDPAR file contains all hospital admissions of Virginia residents between 1984 and 1989. It provides information on date of admission and discharge, up to five International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) diagnostic codes, and up to three ICD-9-CM surgical procedure codes and hospital provider number.

The Medicare demographic files were used to identify enrollment in a health maintenance organization as well as to provide demographic characteristics of the patient. The 1990 Census Bureau Zip Code Level file (Summary Tape 3B) was used to identify age- and race-specific sociodemographic characteristics of the geographic area (Zip code) in which the cases resided.

Sample

The sample consisted of Virginia residents aged 65 years and over who were diagnosed from 1986 though 1989 with breast, colorectal, lung, or prostate cancer and who were treated at a hospital that reported to the VCR. For MEDPAR, incident cases included patients who had a hospital discharge record from MEDPAR with the first occurrence of a specific cancer ICD-9-CM code for the years 1986 through 1989 with no previous mention of this cancer as far back as 1984 (5, 6, 22). Thus, no cancer admissions occurred for a minimum of 2 years prior to this date. VCR and MEDPAR data were linked by social security number, and matches were confirmed by date of birth and sex.

Excluded from analysis were persons diagnosed at autopsy and persons enrolled in a health maintenance organization during a 1-year period subsequent to diagnosis (5, 6). Also excluded from analysis, as mentioned above, were people with cancer who were seen only at hospitals not reporting to VCR.

Variables

Analytic variables were selected to represent demographic, clinical, administrative, and treatment domains.

Demographic variables included age, race, gender, education, income, and geographic region. These variables were available for both VCR and MEDPAR. MEDPAR was chosen as the primary source.

Income and education were derived from the 1990 Census Zip code level file. An income value was assigned to each case as the median income for persons in that Zip code of the same race and age group (65–74, ≥75 years). An education value was assigned as the race-specific proportion of the population in the case's Zip code with less than a ninth-grade education. Dichotomous variables were created for income and education. Cases with an income value greater than the median value of $20,000 were assigned a value of Higher income = 1 (Higher income = 0 if less than median). For education, if the proportion of the Zip code population with less than a ninth-grade education was less than the median value of 10 percent, then More education = 1 (More education = 0 if greater than median). These variables represent the social environment of the subject rather than any person-level value of income or educational attainment (24, 25).

Cases were classified as urban or rural on the basis of a county-level Metropolitan Statistical Area classification used by the Health Care Financing Administration.

Clinical variables included comorbidity and stage. Stage was available only from the VCR and was based
on the four-level summary stage (in situ, local, regional, and distant) used in the SEER study.

To capture overall health status not related to cancer, a comorbidity score was created based on a modified Charlson index (26). The Charlson score is a weighted summary of all prognostically significant medical conditions. For this study, the Dartmouth-Manitoba ICD-9-CM conversion algorithm (27) was used. All ICD-9-CM codes on the MEDPAR record for hospitalizations in the 365 days prior to the month of diagnosis were used to create the score, with the exception of the four cancer site codes. For analysis, a dichotomous variable was created to distinguish between those with no comorbidities (No comorbidity = 1) and those with at least one noncancer comorbidity (No comorbidity = 0).

Administrative variables, available only from MEDPAR, included the number of cancer admissions and the position of the cancer diagnosis on the list of discharge diagnoses. Dichotomous variables were created by considering whether there was only one or more than one cancer admission and whether the cancer diagnosis for the initial visit was in the first or a later position on the hospital claim. The first position is the principal diagnosis and should represent the reason for admission.

Treatment variables used in analysis depended on the data source. Analysis using VCR as the standard examined whether the case had definitive surgical therapy, nonsurgical treatment, or no treatment. Since information on nonsurgical treatments was generally not available on MEDPAR, analysis using MEDPAR as the "gold standard" considered whether the case had definitive surgical therapy, a diagnostic procedure only, or no surgery.

Definitive surgical therapy was defined as nondiagnostic surgical procedures occurring in the initial treatment window (from 2 months prediagnosis to 4 months postdiagnosis for lung and prostate cancers and 6 months postdiagnosis for breast and colon cancers). Definitive surgical therapy did not include biopsy for any site or transurethral resection of the prostate for prostate cancer; these were categorized as diagnostic procedures. Nonsurgical treatment consisted of radiation, hormonal therapy, or chemotherapy. The diagnostic and surgical codes used are included in Appendix 1.

Analysis

The ability of each database to identify incident cases was measured by using the complementary database as the gold standard. For example, when MEDPAR was the standard, this was assessed as the proportion of MEDPAR cancer cases found in VCR. The ability of VCR and MEDPAR to each identify definitive surgical therapy was estimated in the same manner.

The association between demographic, clinical, and treatment variables with MEDPAR missing a case that was captured by the VCR was measured by the odds ratio and 95 percent confidence interval. Odds ratios greater than one signified that cases with the specified value of the variable were more likely to be missed by MEDPAR (i.e., to not match with the VCR). Confidence intervals that did not contain one represented statistically significant relations at the \( \alpha = 0.05 \) level. Treatment comparisons were stratified (i.e., controlled) by stage. Similar analyses were performed to determine the factors related to VCR missing a case identified by MEDPAR.

Capture-recapture techniques (28) were used to estimate the actual cancer population size, based on the concordance and discordance of the data sources. If VCR identifies \( M \) cases and MEDPAR identifies \( n \) cases, \( m \) of which are common to both sources, then the estimated number of cases in the entire population of cases at reporting hospitals will be \( N = [(M + 1) \times (n + 1)]/(m + 1) - 1 \). With this estimate of the population, the sensitivity of each source alone, as well as those of the combined sources, was estimated.

RESULTS

The ability of each database to identify incident cases is shown in Table 1. When the VCR was used as the gold standard, MEDPAR identified 83 percent of breast and lung cancer cases, 81 percent of colorectal cancers, and 73 percent of prostate cancers. Most of the matches (95–98 percent) were within the diagnostic window. Similarly, when MEDPAR was used as the standard, from 70 to 82 percent of cancer cases seen at reporting hospitals were found in the VCR.

Results in Table 2 address the question of whether there is systematic bias in the case missed by

| TABLE 1. VCR* and MEDPAR* matching success for incident cases, 1986-1989 |
|-----------------------------|-----------------------------|-----------------------------|
|                             | Breast | Colorectal | Lung | Prostate |
| No. of cases in VCR         | 3,690  | 4,690      | 5,781 | 4,495    |
| % found in MEDPAR           | 83     | 81         | 83    | 73       |
| No. of cases in MEDPAR      | 3,741  | 4,818      | 6,439 | 4,675    |
| % found in VCR             | 82     | 79         | 74    | 70       |

* VCR, Virginia Cancer Registry, MEDPAR, Medicare Provider Analysis and Review File.
MEDPAR. Demographic variables related to increased risk of being missed by MEDPAR include younger age, male gender, living in an urban area (only significant for breast and prostate cancer), and higher social class. Stage of disease was found to be an important predictor of who will be missed by MEDPAR for all sites but lung. In particular, in situ cases are much more likely to be missed than those with local disease, and those with regional and distant disease are less likely to be missed. In general, patients with no comorbidity are much more likely to be missed by MEDPAR. Patients who receive definitive surgical therapy are much less likely to be missed by MEDPAR, and patients receiving only nonsurgical therapy (chemotherapy or radiation) are more likely than those with no treatment to be missed by MEDPAR.

Table 3 addresses the similar question of whether there is systematic bias in who was missed by VCR. Older age, female gender, and nonwhite race tended to be associated with increased risk of being missed by VCR. The VCR was more likely to miss breast and prostate cancer cases of patients who had comorbid conditions. An increased risk of VCR missing a case was found if there was more than one hospitalization, if the cancer diagnosis was in the second through fifth position on the MEDPAR record, or if no cancer-related surgical procedure was performed (definitive or biopsy). The exception was lung cancer, where cases were more likely to be missed if they had only one cancer admission, and colorectal cancer, where there was increased risk of being missed for cases who had a diagnostic procedure only.

The results in table 4 address the ability of each source to identify definitive surgical therapy. Using surgeries recorded on the VCR as the standard, there were MEDPAR records for 67–80 percent of the surgeries. On the other hand, using MEDPAR as the standard, VCR had 82–87 percent of the surgeries. Finally, the analysis looks at the result of aggregating information from VCR and MEDPAR on estimates of incidence (table 5). Together, VCR and MEDPAR identified 4,374 women with breast cancer at reporting hospitals from 1986 to 1989. Capture-recapture techniques estimate the total number of breast cancer cases at reporting hospitals to be 4,517. Thus, while the sensitivity of the individual sources is 0.82–0.83, the sensitivity of the aggregated information is 0.97 (4,374/4,517). Similarly, for the other cancer sites, the sensitivity of the single sources ranges from 0.70 to 0.83, but increases to 0.92–0.97 when the information from VCR and MEDPAR are aggregated.

### Table 2. Risk of being missed by MEDPAR* when reported by VCR*: demographic, clinical, and treatment factors, 1986–1989

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Breast</th>
<th>Colorectal</th>
<th>Lung</th>
<th>Prostate</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OR</strong></td>
<td><strong>95% CI</strong></td>
<td><strong>OR</strong></td>
<td><strong>95% CI</strong></td>
<td><strong>OR</strong></td>
</tr>
<tr>
<td><strong>Pack</strong></td>
<td><strong>Pack</strong></td>
<td><strong>Pack</strong></td>
<td><strong>Pack</strong></td>
<td><strong>Pack</strong></td>
</tr>
<tr>
<td><strong>Age £74 years</strong></td>
<td>1.08</td>
<td>0.91–1.28</td>
<td>1.64</td>
<td>1.41–1.90</td>
</tr>
<tr>
<td><strong>Male</strong></td>
<td>1.30</td>
<td>1.12–1.50</td>
<td>1.54</td>
<td>1.32–1.80</td>
</tr>
<tr>
<td><strong>White</strong></td>
<td>0.93</td>
<td>0.73–1.18</td>
<td>0.88</td>
<td>0.73–1.07</td>
</tr>
<tr>
<td><strong>Urban</strong></td>
<td>1.28</td>
<td>1.02–1.61</td>
<td>1.04</td>
<td>0.87–1.25</td>
</tr>
<tr>
<td><strong>More education†</strong></td>
<td>1.49</td>
<td>1.25–1.77</td>
<td>1.21</td>
<td>1.04–1.41</td>
</tr>
<tr>
<td><strong>Higher income‡</strong></td>
<td>1.77</td>
<td>1.48–2.12</td>
<td>1.64</td>
<td>1.39–1.92</td>
</tr>
<tr>
<td><strong>Clinical</strong></td>
<td><strong>Pack</strong></td>
<td><strong>Pack</strong></td>
<td><strong>Pack</strong></td>
<td><strong>Pack</strong></td>
</tr>
<tr>
<td><strong>Stage</strong></td>
<td><strong>Pack</strong></td>
<td><strong>Pack</strong></td>
<td><strong>Pack</strong></td>
<td><strong>Pack</strong></td>
</tr>
<tr>
<td><strong>In situ</strong></td>
<td>2.61</td>
<td>2.04–3.33</td>
<td>5.81</td>
<td>4.69–7.20</td>
</tr>
<tr>
<td><strong>Local</strong></td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td>1.00</td>
</tr>
<tr>
<td><strong>Regional</strong></td>
<td>0.66</td>
<td>0.53–0.82</td>
<td>0.50</td>
<td>0.43–0.58</td>
</tr>
<tr>
<td><strong>Distant</strong></td>
<td>1.36</td>
<td>1.01–1.83</td>
<td>0.64</td>
<td>0.52–0.79</td>
</tr>
<tr>
<td><strong>No comorbidity</strong></td>
<td>2.00</td>
<td>1.51–2.64</td>
<td>1.80</td>
<td>1.48–2.21</td>
</tr>
<tr>
<td><strong>Treatment§</strong></td>
<td><strong>Pack</strong></td>
<td><strong>Pack</strong></td>
<td><strong>Pack</strong></td>
<td><strong>Pack</strong></td>
</tr>
<tr>
<td><strong>Initial therapy</strong></td>
<td><strong>Pack</strong></td>
<td><strong>Pack</strong></td>
<td><strong>Pack</strong></td>
<td><strong>Pack</strong></td>
</tr>
<tr>
<td><strong>Definitive surgery</strong></td>
<td>0.26</td>
<td>0.20–0.34</td>
<td>0.60</td>
<td>0.49–0.74</td>
</tr>
<tr>
<td><strong>Chemotherapy/radiation</strong></td>
<td>2.24</td>
<td>1.59–3.14</td>
<td>2.31</td>
<td>1.69–3.14</td>
</tr>
<tr>
<td><strong>No surgery</strong></td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td>1.00</td>
</tr>
</tbody>
</table>

* MEDPAR, Medicare Provider Analysis and Review File; VCR, Virginia Cancer Registry; OR, odds ratio; CI, confidence interval
† Less than 10% without high school for Zip code
‡ More than $20,000 income for Zip code
§ Controlling for stage.
TABLE 3. Risk of being missed by VCR* when reported by MEDPAR*: demographic, clinical, administrative, and treatment factors, 1986-1989

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Breast</th>
<th>Colorectal</th>
<th>Lung</th>
<th>Prostate</th>
<th>Breast</th>
<th>Colorectal</th>
<th>Lung</th>
<th>Prostate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age ≤74 years</td>
<td>0.88</td>
<td>0.99</td>
<td>0.80</td>
<td>0.66</td>
<td>0.75-1.04</td>
<td>0.86-1.14</td>
<td>0.71-0.90</td>
<td>0.58-0.75</td>
</tr>
<tr>
<td>Male</td>
<td>0.78</td>
<td>0.69</td>
<td>0.68</td>
<td>0.89</td>
<td>0.66-0.90</td>
<td>0.58-0.82</td>
<td>0.60-0.78</td>
<td>0.77-1.03</td>
</tr>
<tr>
<td>White</td>
<td>0.53</td>
<td>1.11</td>
<td>1.11</td>
<td>1.03</td>
<td>0.43-0.65</td>
<td>0.93-1.32</td>
<td>0.97-1.28</td>
<td>0.89-1.18</td>
</tr>
<tr>
<td>Urban</td>
<td>0.87</td>
<td>0.97</td>
<td>1.13</td>
<td>1.00</td>
<td>0.82-1.15</td>
<td>0.84-1.12</td>
<td>1.01-1.27</td>
<td>0.88-1.14</td>
</tr>
<tr>
<td>More education†</td>
<td>0.97</td>
<td>0.92</td>
<td>1.01</td>
<td>0.80</td>
<td>0.80-1.17</td>
<td>0.77-1.09</td>
<td>0.89-1.16</td>
<td>0.69-0.95</td>
</tr>
<tr>
<td>Higher income‡</td>
<td>0.97</td>
<td>0.96</td>
<td>1.01</td>
<td>0.80</td>
<td>0.62-0.93</td>
<td>0.81-1.13</td>
<td>0.90-1.13</td>
<td>0.70-0.92</td>
</tr>
</tbody>
</table>

Clinical

| No comorbidity       | 0.76   | 0.73       | 1.68   | 0.77     | 0.62-0.93 | 0.62-0.86  | 1.68-2.11 | 0.67-0.87 |
| One cancer admission§ | 0.46   | 0.20       | 0.30   | 0.28     | 0.37-0.57 | 0.19-0.31  | 0.27-0.34 | 0.24-0.32 |
| Cancer Dx in 1st position† | 0.11 | 0.20       | 0.30   | 0.28     | 0.09-0.13 | 0.17-0.24  | 0.27-0.34 | 0.24-0.32 |

Treatment

| Initial therapy      | 0.09   | 0.14       | 0.34   | 0.18     | 0.07-0.10 | 0.12-0.17  | 0.28-0.41 | 0.14-0.24 |
| Definitive therapy   |        |            |        |          | 0.89     | 0.53-1.50  | 1.70-2.78 | 0.31-0.35 |
| Diagnostic procedure only | 0.00 | 2.17       | 0.43   | 0.31     | 1.00     | 1.00-1.50  | 0.38-0.49 | 0.27-0.35 |
| No surgery           |        |            |        |          | 1.00     | 1.00       | 1.00      | 1.00      |

* VCR, Virginia Cancer Registry; MEDPAR, Medicare Provider Analysis and Review File; OR, odds ratio; Cl, confidence interval
† Less than 10% without high school for Zip code
‡ More than $20,000 income for Zip code.
§ Reference group is more than one admission.
‖ Reference group is 2nd-5th diagnostic code position on MEDPAR record.

TABLE 4. VCR* and MEDPAR* matching success for definitive surgical therapy, 1986-1989

<table>
<thead>
<tr>
<th></th>
<th>Breast</th>
<th>Colorectal</th>
<th>Lung</th>
<th>Prostate</th>
</tr>
</thead>
<tbody>
<tr>
<td>No of surgeries in VCR</td>
<td>3,336</td>
<td>4,118</td>
<td>1,262</td>
<td>670</td>
</tr>
<tr>
<td>% found in MEDPAR</td>
<td>80</td>
<td>77</td>
<td>67</td>
<td>70</td>
</tr>
<tr>
<td>No of surgeries in MEDPAR</td>
<td>3,093</td>
<td>3,836</td>
<td>1,031</td>
<td>543</td>
</tr>
<tr>
<td>% found in VCR</td>
<td>87</td>
<td>83</td>
<td>82</td>
<td>87</td>
</tr>
</tbody>
</table>

* VCR, Virginia Cancer Registry; MEDPAR, Medicare Provider Analysis and Review File

DISCUSSION

This study evaluated the relative ability of two data sources (VCR and MEDPAR) to identify cancer cases and treatment independently. An important difference between this study and that of McBean et al. (5, 6) and Whittle et al. (22) is that the latter compared population rates rather than matching individuals to look for discrepancies. It is possible that rates would be similar because of "compensating errors," with each source identifying different but similarly sized groups of patients. For example, there were 3,690 cases of breast cancer during the study period as identified by VCR and 3,741 found through MEDPAR, yielding virtually identical incidence rates from each source. Yet, these were not all the same cases. Only 83 percent of the breast cancer cases in VCR were found in MEDPAR, while only 82 percent of the possible incident cases in MEDPAR were in the VCR.

Factors representing bias in reporting to MEDPAR included selected demographic characteristics, stage, lack of noncancer comorbid illnesses, and receiving no cancer treatment. Similarly, some demographics, administrative factors, and initial therapy were associated with bias in reporting to VCR.

The implications of bias in reporting to this and other registries is recognition that the entire target population might not be covered and that lack of coverage may not be random. Thus, total incidence may be affected, and reporting bias may cause underrepresentation of the cancer problem in certain population subgroups. Biased misclassification will lead to biased comparisons whose direction cannot be anticipated. Even if there are no systematic differences in identifying cancer cases, comparisons with respect to risk factors will be biased toward the null. These findings should make other researchers aware of possible biases associated with their work with either cancer registries or other secondary databases such as Medicare.

Cases not reported to the VCR were more likely to have their cancer diagnosis found in the second through fifth position in the MEDPAR record rather than in the first. That implies that cancer may not have been the primary reason for the hospitalization, so that MEDPAR may have identified a prevalent rather than an incident case. While the method used here has been used by others (5, 6, 22), some cases termed incident may have been prevalent cases seen as a recurrence after more than 2 years. More complex methods to identify incident cases that look at the position of cancer diagnostic code and the surgical interventions might improve this, but at the expense of missing more cases.

Finally, there was considerable disagreement between MEDPAR and VCR with respect to initial definitive surgical therapy. A reason for surgical treatment recorded only in VCR is missing inpatient MEDPAR claims. Fisher et al. (14) examined the proportion of hip fracture cases identified from part B Medicare data and found that across the United States as a whole 9 percent were identified from Part B alone, and in Virginia the number was 32.8 percent. Whittle et al. (22) found that resection rates for incident cases of cancer were underestimated by Medicare as compared with SEER.

Possible reasons for a surgical treatment recorded in MEDPAR not captured by VCR include nonreporting of the case, hospital miscoding or upcoding of treatment on a Medicare claim (29), and the addition of a surgical treatment to the original treatment plan reported to the VCR.

Cancer registries collect detailed clinical information for incident cases and their treatment. However, because cancer registries have traditionally been hospital based, current trends toward outpatient diagnosis and treatment may reduce their coverage. Using Medicare claims alone for cancer surveillance is insufficient because of limitations in tumor-specific information such as stage and because of the difficulty in differentiating between prevalent and incident cases. However, as a supplement to currently available cancer registries, Medicare, particularly the new Medicare data system, can provide essential data on comorbidity, both for inpatient and outpatient treatments as well as to identify additional incident cases. In this study, the combined data increased the sensitivity from 0.70–0.83 for a single source to 0.92–0.97 for the aggregated sources.

The potential for extending the methodology used in this study to other populations and databases, such as private insurers, needs further exploration. This is particularly important because increasing enrollment in health maintenance organizations, which currently maintain less-specific data, will make future validation studies more difficult. Such studies will have to rely on other data sources, such as statewide hospital discharge files, or work directly with health maintenance organizations to remedy this.

Despite the limitations of secondary data, administrative claims can serve as a relatively inexpensive supplement to tumor registries. Linkages between cancer registries and other data sources may be especially useful for quality control, for assessing the impact of new therapeutic interventions and for judging the dissemination of new technologies and practice guidelines.

ACKNOWLEDGMENTS

Supported by a grant from the Agency for Health Care Policy and Research (AHCPR R01 HS0659–01A1)

REFERENCES


APPENDIX

Diagnostic and Surgical Codes Used in Analysis

ICD-9-CM DIAGNOSTIC CODES

Breast: 174, 174.9; 233
Colorectal: 153–153.9; 154; 154.0; 154.1; 230.3; 230.4
Lung: 162–162.9; 231; 231.2; 231.9
Prostate: 185–185.9; 233.4

ICD-9-CM SURGICAL CODES

Breast:
- Definitive surgical therapy: 85.4–85.48; 85.21–85.23
- Biopsy: 85.11; 85.12; 85.19

Colorectal:
- Definitive surgical therapy: 45.4–45.43; 45.49; 45.7–45.79; 45.8; 48.3–48.35; 48.4–48.49; 48.5; 48.6–48.69
- Biopsy: 45.23–45.29; 48.23–48.29

Lung:
- Definitive surgical therapy: 32.0–32.2; 32.28–32.29
- Biopsy: 33.22–33.29

Prostate:
- Definitive surgical therapy: 60.2–60.69
- Diagnostic procedure: 60.11; 60.12; 60.18; 60.2

VCR-MEDPAR Comparison 233