Comparison of Approaches for Estimating Prevalence Costs of Care for Cancer Patients

What Is the Impact of Data Source?

K. Robin Yabroff, PhD,* Joan L. Warren, PhD,* Jessica Banthin, PhD,† Deborah Schrag, MD,‡ Angela Mariotto, PhD,* William Lawrence, MD, MS,† Angela Meekins, B.S.,§ Marie Topor, B.S.,§ and Martin L. Brown, PhD*

Background: National prevalence costs of medical care can be key inputs in health policy decisions. Cost estimates vary across data sources, patient populations, and methods, however, the objective of this study was to compare 3 approaches for estimating the prevalence costs of colorectal cancer (CRC) care using different data sources, but similar patient populations and methods.

Methods: We identified prevalent CRC patients aged 65 and older from: (1) linked Surveillance Epidemiology and End Results (SEER) registry-Medicare data, (2) Medicare claims only, and (3) the Medical Expenditure Panel Survey (MEPS). Controls were matched by sex, age-group, and geographic location. Mean per person total and net costs, measured as the difference between patients and controls, were compared for each approach during a similar observation period. The SEER-Medicare approach was our reference, and we evaluated the impact of patient selection criteria with sensitivity analyses. Aggregate prevalence estimates were also compared.

Results: We found considerable variability across the different approaches to estimating prevalence costs of CRC. Mean net annual per person estimates in the SEER-Medicare reference were $5341 (95% CI: $5243, $5439), compared with $8736 (95% CI: $8203, $9269) for the Medicare claims only and $11,614 (95% CI: $7566, $15,663) for the MEPS. Aggregate national estimates of net prevalence costs of CRC in 2004 ranged from $4524 million, using the SEER-Medicare approach, to $9629 million, using the MEPS approach. Estimates varied by data source based on the payors included and identification of prevalent CRC patients.

Conclusions: CRC prevalence cost estimates vary substantially depending on the data sources. Our findings have implications for estimating prevalence costs for other cancers and other diseases without registry systems that can be used to identify newly diagnosed individuals as well as those diagnosed less recently.

Key Words: health care costs, health services research, cost and cost analysis, colorectal neoplasms, Medicare, SEER program, MEPS

(Med Care 2009;47: S64–S69)

Aggregate measures of the burden of disease are routinely used to describe the health of populations, establish public health goals, and evaluate allocation of health care resources.1 Disease incidence, prevalence, and mortality, as well as economic measures, such as the cost of medical care are commonly used to quantify the aggregate burden of disease.1–3 In particular, aggregate measures of the economic burden of disease are often reported for a specific calendar year, and are based on the cost of medical care in that year to all individuals diagnosed with or living with that disease. These aggregate costs are also referred to as prevalence costs, because they encompass care delivered to individuals across the disease trajectory, including the newly diagnosed, the long-term survivors, as well as those who are at the end-of-life. These aggregate prevalence costs can be used to inform health policy decisions on the structure of insurance benefits, eligibility criteria for public programs, and budgeting for future program costs.2,3

Prevalence costs have been estimated from a variety of data sources in the United States, including insurance claims, billing systems, hospital discharge databases, and surveys, although few studies have compared cost estimates across data sources. In one of the only studies to compare cancer prevalence costs from different data sources in the United States,4 estimates were found to vary widely.5 In this study, we compared prevalence costs for elderly colorectal cancer (CRC) patients using similar methods in 3 different data sources—cancer registry data linked to Medicare claims, Medicare claims alone, and the Medical Expenditure Panel Survey (MEPS). Because CRC is a common cancer and
primarily a disease of the elderly,6 our sample populations are ideal for this comparison. Our findings may have implications for the estimation and interpretation of prevalence costs in other disease areas, which do not have the advantage of registry systems that can be used to identify newly diagnosed patients, as well as long term survivors of disease.

**METHODS**

**Data Sources**

We used 3 data sources to compare approaches for estimating prevalence costs of CRC care: (1) Surveillance Epidemiology and End Results (SEER) program registry data linked to Medicare claims data (SEER-Medicare), (2) Medicare claims only, and (3) the MEPS. We selected the SEER-Medicare approach as our reference because it was likely to be the most accurate in identifying CRC patients.

**Linked SEER-Medicare Data**

As described in the companion paper on incidence costs,7 the SEER registries collect information about all incident cancer patients from geographically defined areas, approximately 14% of the US population during the years of our study.6 Among individuals aged 65 and older with a cancer diagnosis in the SEER data, 94% have been linked with Medicare enrollment data.8 The NCI has also created a data file that identifies a 5% random sample of all Medicare beneficiaries residing in SEER areas and an indicator of a cancer diagnosis listed in SEER. Beneficiaries with cancer were removed from the 5% sample and the remaining beneficiaries without cancer were potential controls. Medicare fee-for-service data include longitudinal claims for covered health care services, including hospital, physician, outpatient, home health, and hospice bills from the time of a person’s Medicare eligibility until death.

**Medicare Claims Only**

The entire 5% random sample of Medicare beneficiaries residing in SEER areas (both with and without cancer) was used as a data source for the claims-only approach.

**MEPS**

The MEPS Household Component is a nationally representative household survey of health care utilization and expenditures for the US noninstitutionalized civilian population.9 The MEPS includes information about health care services covered by all sources of payment, including Medicare, private insurance, other public payors such as Medicaid and the Veterans Health Administration, and out of pocket payments. During the period of our study, survey response rates ranged from 61% to 71%.10

**Study Populations**

Within each data source, we identified prevalent CRC patients aged 65 and older using standard definitions for each data source and then matched noncancer controls by sex, age-group, and geographic location during the observation period (Table 1).

**Linked SEER-Medicare**

A total of 73,050 CRC patients were identified from registry data and 135,814 noncancer controls were identified from the sample of Medicare beneficiaries without cancer living in SEER areas.

**Medicare Claims Only**

CRC patients were identified by either having an inpatient claim with a CRC diagnosis code or 2 outpatient claims with CRC diagnosis codes at least 60 days apart, but within 365 days. This algorithm is similar to those used in other settings, such as the CMS Chronic Condition Data Warehouse.3 The final sample included 3575 CRC patients and 17,875 noncancer controls (as defined by the diagnostic code algorithm).

**MEPS**

A total of 196 CRC patients were identified because they reported CRC as bothering them, leading to medical care (eg, physician visit), or resulting in not being able to perform usual activities, including work, school, and housework. Noncancer controls did not report any cancer and were frequency matched to patients (N = 12,152).

**Analyses**

Mean per person annual costs of care were calculated for CRC patients and controls. Mean net per person annual costs of care were calculated as the difference in costs between CRC patients and matched controls, and reflect the costs associated with CRC. For the SEER-Medicare and Medicare claims only approaches, payments, rather than billed charges were used as a proxy for medical care costs. Medicare payments are derived from reimbursement formulas intended to reflect the average resource utilization for a specific service, whereas charges reflect price-setting rather than resource consumption, and as a result, are thought to be a poor proxy of the true economic cost of medical care.11 Mean per person annual and net expenditures from the MEPS data were generated by applying sample weights that account for sampling probability and adjust for potential nonresponse bias. All estimates are reported in 2004 dollars.

**Sensitivity Analysis**

Our base analysis focused on patients with CRC as their only cancer. We conducted sensitivity analyses to evaluate the impact of (1) including CRC patients with other cancers and (2) using more years of claims (1986–2002 vs. 1996–2002) in the Medicare claims only approach to identify CRC patients and estimate costs during the observation period.

**Aggregate CRC Prevalence Cost Estimates**

Mean annual, per-person prevalence cost estimates are often used to calculate aggregate national cost estimates for a calendar year. We used complete CRC prevalence in the US among individuals aged 65 and older as of December 31, 2004, approximately 829,068 men and women.6,12 The complete prevalence estimate represents individuals alive on a specific date who ever had a history of cancer,6 and is a standard measure of chronic disease prevalence. We then multiplied the
mean annual per person cost estimate (both total and net) by CRC prevalence in the US for each approach.

**RESULTS**

Mean age during the observation period was similar across approaches (Table 2). Survival following patient diagnosis or identification varied substantially between the SEER-Medicare and Medicare claims only approaches, reflecting differences in the number of years used to identify CRC patients and controls.

**Mean Per Person Total and Net Annual Costs**

Mean per person total and net annual estimates varied across approach (Table 3). Mean total annual per person estimates were $12,231 (95% confidence interval [CI]: $12,188, $12,274) in the SEER-Medicare reference, $17,579 (95% CI: $17,073, $18,086) in the Medicare claims only, and $18,359 (95% CI: $14,320, $22,398) in the MEPS. Annual per person payments for CRC patients and matched controls are listed in Figure 1 by payor type, including Medicare, private insurance, Medicaid, out-of-pocket, and other public and private payors. Although the Medicare payments are similar in the MEPS and SEER-Medicare approaches for CRC patients, Medicare payments for matched controls were much lower in MEPS than in the other approaches (Fig. 1). As a result, net estimates vary across approaches.

Mean annual per person net estimates were $5457 (95% CI: $5362, $5552) in the SEER-Medicare reference, $8736 (95% CI: $8203, $9269) in the Medicare claims only, and $11,614 (95% CI: $7566, $15,663) in the MEPS.

**Sensitivity Analysis**

In the sensitivity analysis including CRC patients with prior cancer diagnoses, net estimates were similar to the base estimate for each approach. Although differences in the MEPS estimates were larger (about 19%), CIs were wide, and overlapped. Notably, when more years of claims were used to identify CRC patients in the Medicare claims only approach, net payments during the same observation period (1998–2002) were $5413 (95% CI: $5066, $5760). This estimate is 38% lower than the Medicare claims only base estimate, but less than 1% different from the SEER-Medicare reference.

**Aggregate Prevalence Cost Estimates**

In 2004, aggregate net national prevalence estimates (Table 4) were approximately $4524 (95% CI: $4445, $4603)
million based on the SEER-Medicare reference, $7243 (95% CI: $6800, $7684) million in the claims only approach, and $9629 (95% CI: $6273, $12,986) million in the MEPS approach.

**DISCUSSION**

In this study, we used 3 sources of data, linked SEER-Medicare, Medicare claims alone, and the MEPS, and similar methods to compare estimates of the prevalence costs of care in elderly CRC patients. We found significant variation in mean per-person estimates across data sources. National aggregate estimates of net CRC prevalence costs in the elderly in the US in 2004 also varied substantially with a range of $4524 million with the SEER-Medicare approach to $9629 million with the MEPS approach. Our goal in this study was to better understand how CRC prevalence cost estimates are likely to be affected by the underlying data source.

There is no gold-standard data source for estimating the prevalence costs of cancer care. Key dimensions of the underlying data sources include national representativeness, completeness of the payment or expenditure data, number of individuals with the condition under study, and accurate identification of both longer-term survivors as well as patients actively receiving care. Identification of longer-term survivors is important because longitudinal studies have shown that costs and health limitations in cancer patients are higher than in similar individuals without cancer, even outside of the initial diagnosis and end of life periods. These dimensions are sources of variation in our estimates, and their relative importance will vary based on the expected use of the estimate and the specific cancer being evaluated.

Both the SEER-Medicare and MEPS approaches have important strengths, but differ with respect to the represen-
The advantages of using the Medicare claims only approaches to estimate prevalence costs for CRC and other cancers are limited, because these data do not reflect all payors and claims algorithms will disproportionately select cancer patients actively receiving cancer care (treated prevalence only). Longer-term survivors not receiving cancer care will be under-represented in the sample because they cannot be identified, even with more years of claims. Thus, claims only approaches will likely overstate prevalence costs of cancer care. The utility of claims only approaches for estimating prevalence costs in diseases other than cancer is unknown.

In a companion manuscript, we evaluated approaches to estimating CRC incidence costs. Differences between incidence and prevalence estimates will likely be greatest for cancers with longer survival, because of the longitudinal

**TABLE 3.** Comparison of Approaches for Estimating Mean Per Person Annual Prevalence Costs of Care for Colorectal Cancer Patients

<table>
<thead>
<tr>
<th></th>
<th>Total Per Person Costs</th>
<th>Net Per Person Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est ($)</td>
<td>95% CI ($)</td>
</tr>
<tr>
<td>SEER-Medicare</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base</td>
<td>12,231</td>
<td>12,188</td>
</tr>
<tr>
<td>Sensitivity*–other non-CRC cancers included</td>
<td>12,803</td>
<td>12,731</td>
</tr>
<tr>
<td>Medicare claims only</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base</td>
<td>17,579</td>
<td>17,073</td>
</tr>
<tr>
<td>Sensitivity*–other non-CRC cancers included</td>
<td>16,643</td>
<td>16,273</td>
</tr>
<tr>
<td>Sensitivity*–more yrs of claims to identify CRC</td>
<td>13,891</td>
<td>13,563</td>
</tr>
<tr>
<td>MEPS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base</td>
<td>18,359</td>
<td>14,320</td>
</tr>
<tr>
<td>Sensitivity*–other non-CRC cancers included</td>
<td>20,491</td>
<td>16,853</td>
</tr>
</tbody>
</table>

*Sensitivity analyses were conducted to assess the impact of including (rather than excluding) patients with other cancer diagnosis, and including additional years of claims to identify colorectal cancer patients and estimating costs for the same observation period (1998–2002).

All estimates in 2004 dollars.
u-shaped cost curve following diagnosis. If survival following diagnosis is short, or the longitudinal cost curve is flatter, prevalence and incidence cost estimates may be more similar. Incidence estimates are particularly useful for cost-effectiveness analyses, whereas prevalence estimates that reflect national spending for disease in a specific year are more useful in policy and coverage decision-making. For diseases where the longitudinal cost curve is flat and incidence estimates are not available, prevalence estimates may be a reasonable substitute for incidence estimates in cost effectiveness analyses.

There were several limitations that affect our ability to directly compare prevalence cost estimates. Notably the MEPS included all payors and all components of medical expenditures, whereas the other approaches were limited to Medicare payments for covered services. Although we attempted to separate estimates by payor type, we can not make direct comparisons of patients identified in the 3 data sources. Our Medicare cost estimates are based only on Medicare payments, and only include the approximately 85% of Medicare enrollees in fee-for-service plans.

In conclusion, CRC prevalence cost estimates vary substantially depending on the data sources, reflecting differences in the payors included, patient selection, and proportion of long-term survivors in each sample. Our findings have implications for estimating prevalence costs for other cancers and other diseases without registry systems that can be used to identify newly diagnosed individuals as well as those diagnosed less recently.

**REFERENCES**


---


<table>
<thead>
<tr>
<th></th>
<th>Total Costs</th>
<th>Net Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate ($)</td>
<td>95% CI ($)</td>
</tr>
<tr>
<td>SEER-Medicare</td>
<td>10,140</td>
<td>(10,105; 10,176)</td>
</tr>
<tr>
<td>5% Medicare claims</td>
<td>14,574</td>
<td>(14,155; 14,995)</td>
</tr>
<tr>
<td>MEPS</td>
<td>15,221</td>
<td>(11,872; 18,569)</td>
</tr>
</tbody>
</table>

*The number of elderly colorectal cancer survivors alive December 31, 2004 estimated to be 829,068.
All estimates in 2004 dollars.