Comparing Cost-of-Illness Estimates from Alternative Approaches: An Application to Diabetes

Amanda A. Honeycutt, Joel E. Segel, Thomas J. Hoerger, and Eric A. Finkelstein

Objective. To compare disease cost estimates from two commonly used approaches.

Data Source. Pooled Medical Expenditure Panel Survey (MEPS) data for 1998–2003.

Study Design. We compared regression-based (RB) and attributable fraction (AF) approaches for estimating disease-attributable costs with an application to diabetes. The RB approach used results from econometric models of disease costs, while the AF approach used epidemiologic formulas for diabetes-attributable fractions combined with the total costs for seven conditions that result from diabetes.

Data Extraction. We used *SAS* version 9.1 to create a dataset that combined data from six consecutive years of MEPS.

Principal Findings. The RB approach produced higher estimates of diabetes-attributable medical spending (\$52.9 billion in 2004 dollars) than the AF approach (\$37.1 billion in 2004 dollars). RB model estimates may in part be higher because of the challenges of implementing the two approaches in a similar manner, but may also be higher because they capture the costs of increased treatment intensity for those with the disease.

Conclusions. We recommend using the RB approach for estimating disease costs whenever individual-level data on health care spending are available and when the presence of the disease affects treatment costs for other conditions, as in the case of diabetes.

Key Words. Cost-of-illness, diabetes, econometrics, attributable fractions

Cost-of-illness (COI) studies are increasingly used to quantify the public health burden associated with a disease, illness, injury, or risk factor (e.g., Gross et al. 1999; American Diabetes Association [ADA] 2003; Fishman et al. 2003; Honeycutt et al. 2003; Finkelstein et al. 2005). These studies estimate the costs associated with a disease, including direct medical costs for diagnosing and treating the disease and indirect costs, such as productivity losses. Quantifying the economic and public health burden of a disease is useful for understanding the impact of one disease relative to others and for establishing priorities for

disease treatment and prevention (Rice 1994). In some cases, COI estimates are broken down to show the distribution of disease costs across payers (e.g., Finkelstein, Fiebelkorn, and Wang 2003), which can help demonstrate the burden borne by specific stakeholders.

COI studies often attempt to estimate the disease-attributable costs that could be avoided if a case of the disease were prevented. Some COI analyses estimate annual costs for the prevalent population, whereas others estimate lifetime costs for the incident population. The two main approaches for estimating prevalence-based disease-attributable costs, the focus of this analysis, are (1) a regression-based (RB) approach applied to individual-level cost data and (2) an attributable fraction (AF) approach applied to aggregate cost data (Miller, Ernst, and Collin 1999). Although we focus on costs attributable to disease, the same approaches can be applied to estimate the costs of risk factors, such as obesity.

The RB approach uses regression analysis to estimate models of medical spending. These models include an indicator variable for the disease of interest and control for individual-level characteristics, such as sociodemographic variables and comorbidities. The coefficient estimates from these models are used to predict individual-level health care spending in the disease population and then to predict health care spending for these individuals if the disease were eliminated (i.e., treating the disease indicator variable as equal to zero). The mean of the difference between these two predicted values provides an estimate of per-person medical spending attributable to the disease.

The AF approach involves identifying the medical conditions that are caused by the disease of interest and obtaining estimates of the aggregate cost of each condition. AFs are then calculated using epidemiologic formulas for each condition. The AFs represent the portions of disease prevalence that are caused by the presence of the disease. Disease-attributable costs are estimated by multiplying the AFs by the aggregate cost of each condition, and then summing across all conditions.

The RB and AF approaches have been widely used to estimate costs attributable to disease or risk factors (e.g., Hodgson and Cohen 1999; ADA 2003; Finkelstein, Fiebelkorn, and Wang 2003). The choice of an approach

Address correspondence to Amanda A. Honeycutt, Ph.D., RTI International and RTI-UNC Center of Excellence in Health Promotion Economics, 118 Beaufain Street, Charleston, SC 29401; e-mail: honeycutt@rti.org. Joel E. Segel, B.A., Thomas J. Hoerger, Ph.D., and Eric A. Finkelstein, Ph.D., are with RTI International and RTI-UNC Center of Excellence in Health Promotion Economics, RTP, NC.

often depends on the available data and how the cost estimates will be used. The RB approach uses individual-level data on health care spending and the presence of disease to assign costs based on a comparison of actual medical spending among people with and without the disease. It therefore captures differences in disease-attributable spending that may be caused by an increased number of health care visits or longer lengths of stay for people with the disease or by higher costs for any particular visit.

The AF approach often includes only diseases known to result from the condition of interest. For example, when disease cost estimates are used to establish damages in legal proceedings, it may be important to clearly define which conditions are attributed to the disease. Disease cost estimates prepared in support of tobacco litigation in the late 1990s frequently used an AF approach, because analysts could specify that costs include only those diseases known to be caused by smoking. However, several recent AF analyses have included the cost of general medical conditions (e.g., Miller, Ernst, and Collin 1999; Collier, Harrison, and McInnes 2002; ADA 2003), which allows for the possibility that the disease leads to higher overall medical spending and raises costs to treat conditions not generally thought to be caused by the disease.

Even if the cost of general medical conditions is included, the AF approach may produce lower estimates than the RB approach, because it uses aggregate disease cost data. In the case of diabetes, using aggregate data to estimate medical costs implicitly assumes that treating a nondiabetes event (e.g., cardiovascular disease [CVD]) costs the same for a person with diabetes as it does for a similar person without diabetes. But the person with diabetes might require a longer hospital stay for the nondiabetes event, because diabetes complicates treatment and raises treatment costs. By not accounting for differences in treatment intensity, the AF approach may underestimate attributable costs.

In this study, we use both RB and AF approaches to estimate diabetes-attributable medical spending.

METHODS

Data

Study data were drawn from the Medical Expenditure Panel Survey (MEPS) for 1998 through 2003. MEPS is a nationally representative survey of the U.S. civilian, noninstitutionalized population administered by the Agency for Healthcare Research and Quality (AHRQ). Household respondents provided

demographic information, self-reported medical conditions, and medical expenditure and utilization information for medical events. For some individuals, self-reported medical expenditures are supplemented with information from medical providers and insurers. MEPS uses a complex survey design and contains population weights to create nationally representative estimates (Agency for Healthcare Research and Quality 2000).

To ensure sufficient sample size for our analysis, we pooled 6 years of data from the consolidated, medical conditions, and individual event files (consisting of office-based visits, hospital inpatient stays, outpatient department visits, emergency room visits, prescribed medicines, and home health files). Because MEPS is an overlapping panel survey, many individuals are in the sample for two consecutive years; thus, samples from year to year are not completely independent. To pool multiple years of MEPS data, we used the approach recommended by AHRQ, which involves applying person-level weights for each year and performing all analyses using survey commands in *Stata* 9.2 to account for the complex survey design (Agency for Healthcare Research and Quality 2006). We excluded individuals who did not have a complete year's worth of data for any year, those with missing age or sample weights, and women who were pregnant at any time during the year. Our final sample contained 162,648 observations on 96,873 unique individuals. We adjusted expenditures to 2004 dollars using the medical care component of the Consumer Price Index (U.S. Department of Labor 2007).

We identified people with diagnosed diabetes mellitus as those who had a MEPS clinical classification code for diabetes (049, 050). Respondents could report having diabetes for a specific medical event (e.g., for a hospital inpatient stay) or as a medical condition not linked to a specific event. Our final sample contained 8,429 observations for people with diabetes (5,289 unique individuals).

RB Approach

We estimated econometric models of annual health care spending and used the resulting coefficient estimates to calculate the marginal effect of diabetes on per-person medical spending. Several approaches have been recommended for modeling health care expenditures (e.g., Manning et al. 1987; Manning 1998; Mullahy 1998; Manning and Mullahy 2001; Buntin and Zaslavsky 2004). We used model selection criteria recommended by Manning and Mullahy (2001) and Buntin and Zaslavsky (2004), including tests for heteroscedasticity and kurtosis in the residuals from alternative functional forms, and found that the most appropriate model for our MEPS sample was a two-part

generalized linear model (GLM) with a gamma distribution and a log link. We used a logit model to predict the probability of having any medical spending

$$P(y_i > 0) = \frac{e^{x_i B}}{1 + e^{x_i B}} \quad (1)$$

We then used a GLM model with a gamma distribution and a log link to estimate the level of expenditures, given positive spending

$$E(y_i | x_i) = e^{x_i B} \quad (2)$$

where y_i represents total medical spending in dollars for individual i , conditional on having positive expenditures. In both parts, x_i includes the continuous variables age and age-squared and indicator variables for sex, race/ethnicity, education, region, rural status, income, no health insurance, the presence of diabetes, and the presence of 10 other conditions that are likely to be correlated with, but not caused by, diabetes (i.e., cancer, injuries, pneumonia, asthma, cardiopulmonary disease, depression, other mental health and substance abuse problems, arthritis, back disorders, and skin disorders). We did not include diabetes-attributable conditions (e.g., CVD, renal disease). Their inclusion would create downward bias in our estimates of diabetes spending, because the portion of diabetes spending that is attributable to CVD or renal disease would then be allocated to those conditions and not to diabetes (Lee, Meyer, and Clouse 2001).

We estimated separate models of equations (1) and (2) for each age group (i.e., <45, 45–64, 65+ years) and used the MEPS sampling weights to produce nationally representative estimates for the civilian, noninstitutionalized population. Although these models were stratified by age, they also controlled for continuous 1-year age increments within each age group. We used the coefficient estimates from equations (1) and (2) to predict total medical spending for the subsample with diabetes. Using coefficient estimates from equation (1), we estimated the predicted probability of positive medical spending for each person. We multiplied this predicted probability by an estimate from equation (2) of the predicted spending level, given positive expenditures. For this same diabetes subsample, we again predicted medical spending, assuming each individual did not have diabetes (i.e., Diabetes = 0). Our estimates of diabetes expenditures are age group-specific means of the difference between the two predictions (i.e., predicted spending for each person with diabetes minus predicted spending if the person did not have diabetes). We estimated bootstrapped 95 percent confidence intervals

around the estimates using 1,000 iterations and accounting for the complex survey design.

AF Approach

We applied the AF approach to estimate diabetes costs by identifying the conditions that are attributable to diabetes and, for each condition, estimating the amount of medical spending caused by diabetes. Our diabetes-attributable spending estimates were calculated as the estimated AF for each condition—the fraction of the prevalence of the condition that is attributable to diabetes—multiplied by total spending for that condition. We then summed across all conditions to estimate total diabetes-attributable spending. We estimated bootstrapped 95 percent confidence intervals using an approach identical to that used for the RB model.

We identified the following conditions as attributable to diabetes: CVD, renal disease, visual disorders, neurological disease, peripheral vascular disease (PVD), metabolic and endocrine system disorders, and Alzheimer's disease. ADA (2003) included all of these, except Alzheimer's disease, in its analysis of diabetes costs. Thacker et al. (2005) found evidence that diabetes contributes to kidney disease, CVD, blindness or visual impairment, Alzheimer's disease, and perinatal conditions in pregnancy. We included Alzheimer's disease as a diabetes-attributable condition, but because of difficulty distinguishing between preexisting and gestational diabetes in MEPS, perinatal conditions were not included. We used three-digit ICD-9 codes to identify each attributable condition in the MEPS sample (Appendix S1). We also included a category for general medical conditions that consisted of all other conditions.

To estimate the diabetes AF for each condition, we used data from the MEPS conditions file. Because the prevalence of diabetes and its attributable conditions increase with age, we estimated AFs for three age groups using the adjusted AF equation. Rockhill, Newman, and Weinberg (1998) and Flegal, Graubard, and Williamson (2004) describe that, in the presence of confounding factors and/or effect modification, the correct formula for calculating the AF is

$$AF_j = pd_j \left[\frac{RR_j - 1}{RR_j} \right] \quad (3)$$

In equation (3), pd_j represents the age-group adjusted prevalence of diabetes in the subsample with the attributable condition j . RR_j denotes the age-adjusted relative risk (RR) of condition j in the diabetes subsample relative to the non-diabetes subsample. We estimated pd_j and RR_j separately for three age groups:

younger than 45 years, 45–64 years, and 65 years and older. Within each age group stratification, our models controlled for sex, race/ethnicity, and age (Benichou 2001; McNutt et al. 2003).

Medical spending estimates for attributable conditions are frequently drawn from the literature or from national health expenditure accounts (Heffler et al. 2005). To limit the possibility that differences between the RB and AF cost estimates result from differences in the underlying data sources, we estimated aggregate spending for each condition using the six MEPS event files (office-based, inpatient, outpatient, emergency room, prescription medicines, and home health). We used a simple algorithm (Appendix S2) to assign costs to diabetes or one of its attributable conditions, using the diagnosis codes and expenditures reported for each event. We estimated total costs for each attributable condition as the population-weighted sum across the three age groups.

RESULTS

The diabetes subsample consisted of 8,429 observations and contained similar fractions of women (50–52 percent), Hispanics (12–13 percent), and blacks (10–11 percent) as the subsample of people without diabetes (Table 1). However, people in the diabetes subsample were significantly older than people in the no-diabetes subsample (60 versus 35 years), were less likely to be uninsured—likely due to their age, and were more likely to have CVD, renal disease, and/or vision disorders.

RB Models

Results from the two-part GLM models are shown by age group and by model stage in Appendix S3. We used results from these age-group-specific models to calculate annual medical spending attributable to diabetes for the diabetes subsample. Individual level and aggregate results are shown in 2004 dollars in Table 2.

Per-person spending attributable to diabetes is considerably higher for the oldest age group than for the 45–64 age group but only slightly higher than spending for the youngest age group. For the 65 and older age group, per-person diabetes-attributable medical spending is \$4,690 per year. Predicted annual per-person spending is \$3,720 for the 45–64 age group and \$4,520 for the age group younger than 45. The relatively high costs for those younger than 45 years may reflect a duration effect if individuals with Type 1 diabetes are disproportionately represented in this age group.

Table 1: Descriptive Statistics—Sample Weighted Means and 95% Confidence Intervals

	<i>Diabetes Sample</i>	<i>No Diabetes Sample</i>
<i>N</i>	8,429	154,219
Age < 45	1,231	105,315
Age 45–64	3,795	33,209
Age 65+	3,403	15,695
Age	60.1	35.2
	(59.6–60.7)	(34.8–35.5)
Female (%)	52.0	49.9
	(50.3–53.6)	(49.5–50.2)
Black (%)	10.9	9.5
	(9.5–12.3)	(8.5–10.4)
Hispanic (%)	12.3	12.6
	(10.5–14.1)	(11.3–14.0)
Uninsured (%)	6.5	11.8
	(5.6–7.3)	(11.3–12.4)
Cardiovascular disease (%)	61.9	13.3
	(60.4–63.4)	(12.9–13.7)
Renal disease (%)	9.8	3.6
	(8.8–10.8)	(3.5–3.8)
Visual disorders (%)	13.5	3.2
	(12.4–14.6)	(3.0–3.3)

We found that aggregate medical spending attributable to diabetes is \$52.9 billion per year. Almost half of these expenditures are for people aged 65 and older, whereas 15 percent are for people younger than age 45.

Table 2: Estimated Aggregate and Per-Person Costs* by Age Group Using the Regression-Based Approach

<i>Age Group</i>	<i>Regression-Based Approach</i>	
	<i>Total Diabetes-Attributable Costs (95% CI)</i>	<i>Per-Person Costs</i>
Under 45	\$8.0 billion (5.7–11.1 billion)	\$4,520 (3,280–6,030)
45–64	\$20.2 billion (17.6–23.2 billion)	\$3,720 (3,320–4,160)
65+	\$24.6 billion (20.9–28.2 billion)	\$4,690 (4,080–5,310)
All ages	\$52.9 billion (47.2–58.8 billion)	\$4,240 (3,870–4,650)

*2004 dollars.

CI, confidence interval.

AF Models

We first estimated the annual aggregate cost of each condition identified as attributable to diabetes (Table 3). The results suggest that medical spending for diabetes alone (i.e., not including the cost of attributable conditions) was \$21.9 billion per year in 2004 dollars. This estimate reflects the costs for events with diabetes as the only listed diagnosis and a portion of the costs for events with diabetes and one or more of its attributable conditions listed.

Estimated CVD spending was \$85 billion per year; endocrine and metabolic disorders and neurological disorders had annual expenditures of \$19.3 and \$16.3 billion, respectively. Total annual costs for general medical conditions were \$426.9 billion. Our estimates suggest that, on average, total medical spending per year for the categories of services included in MEPS was \$607.9 billion in 2004 dollars.

Table 3 also shows estimates of diabetes AFs. For all of the attributable conditions, the AFs are somewhat higher for those in the 45–64 years age group than for those in the youngest and oldest age groups. We also found that 1 percent or less of the prevalence of general medical conditions is attributable to diabetes.

We multiplied AFs by the total annual costs for each attributable condition to generate estimates of diabetes-attributable costs. These are shown by condition and age group in the last column of Table 3. Our estimates attribute \$6.9 billion of CVD costs (8.2 percent) to diabetes. Of the \$19.3 billion in annual costs for endocrine and metabolic disorders, \$1.81 billion are attributed to diabetes (9.4 percent); of the \$16.3 billion for neurological disorders, \$1.37 billion are attributed to diabetes (8.4 percent). Although the estimated AFs for general medical conditions are low, \$2.59 billion of general medical conditions costs are attributed to diabetes because of large annual costs for these conditions.

Table 4 shows aggregate and per-person annual medical expenditures attributable to diabetes based on the AF approach. These results suggest that \$37.1 billion in medical spending per year is attributable to diabetes. Per-person spending estimates vary by age, but are similar for people in the 45–64 and 65 and older age groups. For these groups, we found per-person diabetes-attributable spending of \$3,020–\$3,090. For the age group younger than 45 years, we estimated per-person annual expenditures of \$2,500.

Comparison of RB and AF Model Results

Results from the RB approach are 43 percent higher than those from the AF approach (Table 4). The RB approach produced significantly higher spending

Table 3: Estimated Attributable Fractions and Annual Medical Expenditures* for Eight Diabetes-Attributable Conditions

	<i>Age Group</i>	<i>Diabetes-Attributable Fraction</i>	<i>Total Annual Condition Expenditures (in Billions)</i>	<i>Diabetes Annual Attributable Expenditures (in Billions)</i>
Diabetes	Under 45	1	\$3.33	\$3.33
	45–64	1	\$9.21	\$9.21
	65+	1	\$9.39	\$9.39
	All ages	1	\$21.93	\$21.93
CVD	Under 45	0.086	\$6.12	\$0.53
	45–64	0.114	\$31.23	\$3.57
	65+	0.060	\$47.62	\$2.84
	All ages	0.082	\$84.97	\$6.94
Renal disease	Under 45	0.015	\$4.88	\$0.07
	45–64	0.089	\$4.68	\$0.42
	65+	0.049	\$4.76	\$0.23
	All ages	0.050	\$14.31	\$0.72
Visual disorders	Under 45	0.052	\$0.96	\$0.05
	45–64	0.148	\$2.64	\$0.39
	65+	0.021	\$8.75	\$0.18
	All ages	0.050	\$12.35	\$0.62
Neurological disease	Under 45	0.028	\$2.32	\$0.07
	45–64	0.101	\$5.22	\$0.53
	65+	0.089	\$8.73	\$0.78
	All ages	0.084	\$16.26	\$1.37
Peripheral vascular disease	Under 45	0.062	\$0.96	\$0.06
	45–64	0.106	\$4.05	\$0.43
	65+	0.097	\$5.92	\$0.58
	All ages	0.098	\$10.93	\$1.07
Metabolic/endocrine system disorders	Under 45	0.085	\$1.94	\$0.16
	45–64	0.107	\$8.59	\$0.92
	65+	0.083	\$8.76	\$0.73
	All ages	0.094	\$19.29	\$1.81
Alzheimer's disease	Under 45	—		
	45–64	0.275	\$0.01	\$0.002
	65+	0.053	\$0.99	\$0.05
	All ages	0.051	\$0.99	\$0.05
GMCs	Under 45	0.001	\$167.70	\$0.16
	45–64	0.009	\$143.50	\$1.31
	65+	0.010	\$115.70	\$1.12
	All ages	0.006	\$426.90	\$2.59
Total diabetes-attributable spending (all ages)				\$37.10

*2004 dollars.

CVD, cardiovascular disease; GMCs, general medical conditions.

Table 4: Estimated Aggregate and Per-Person Medical Expenditures* by Age Group and by Costing Approach

<i>Age Group</i>	<i>Regression-Based Approach</i>		<i>Attributable Fraction Approach[†]</i>	
	<i>Total Diabetes-Attributable Spending (95% CI)</i>	<i>Per-Person Spending (95% CI)</i>	<i>Total Diabetes-Attributable Spending (95% CI)</i>	<i>Per-Person Spending (95% CI)</i>
Under 45	\$8.0 billion (5.7–11.1 billion)	\$4,520 (3,280–6,030)	\$4.4 billion (3.7–5.3 billion)	\$2,500 (2,130–2,930)
45–64	\$20.2 billion (17.6–23.2 billion)	\$3,720 (3,320–4,160)	\$16.8 billion (15.3–18.4 billion)	\$3,090 (2,920–3,250)
65+	\$24.6 billion (20.9–28.2 billion)	\$4,690 (4,080–5,310)	\$15.9 billion (14.4–17.4 billion)	\$3,020 (2,840–3,190)
Total	\$52.9 billion (47.2–58.8 billion)	\$4,240 (3,870–4,650)	\$37.1 billion (34.3–40.2 billion)	\$2,980 (2,850–3,110)

*2004 dollars.

[†]Includes costs from general medical conditions.

estimates than the AF approach for age groups younger than 45 and older than 64 years. Per-person spending estimates were also higher from the RB approach—most notably for the youngest age group, for whom the RB estimates are \$4,520 per person versus \$2,500 from the AF approach.

Results from Alternative Specifications

To better understand why the RB and AF approaches produced such different cost estimates, we estimated several alternative specifications of both models. We first examined the impact of using different sets of control variables in the RB model. One specification included only age, sex, and race/ethnicity as independent variables, which led to an estimate of \$54.7 billion, nearly 3.5 percent higher than our baseline estimate of \$52.9 billion. Because this model does not fully control for observed differences between the diabetes and non-diabetes populations, it is not surprising that the resulting estimates are higher. For another specification, we included diabetes-attributable conditions, such as CVD, renal disease, and vision impairment, in addition to the other independent variables in our baseline model. As expected, this specification lowered the estimates for diabetes because it “overcontrolled” by attributing some costs to attributable conditions (e.g., CVD), that are legitimately attributable to diabetes. This specification resulted in an estimate of \$49 billion—approximately 7 percent lower than our baseline RB estimate. Another

specification included risk factors for diabetes—obesity ($\text{BMI} > 29$) and hypertension—in addition to the variables from the previous model. Because obesity measures were available only for the 2000–2003 period, we estimated the percentage reduction in costs between this alternative specification and our baseline model re-estimated for this time period, which resulted in estimates that were 18 percent lower than baseline costs, or approximately \$43.3 billion. However, it too likely overcontrols because some diabetes costs are assigned to obesity, hypertension, CVD, and other attributable conditions.

We also considered the effect of alternative specifications of the AF model. First, we analyzed the impact of using a broader range of ICD-9 codes to define CVD: 390–459, as used by the Centers for Disease Control and Prevention (CDC), and 390–459 and 745–747, as used by the American Heart Association (AHA). These broader definitions led to a decrease in estimated costs of 2 percent (to \$36.3 billion) and 8 percent (to \$34 billion) for the CDC and AHA definitions, respectively. The lower costs were a result of lower CVD AFs than in our baseline model, due to a weaker link between the added conditions and diabetes.

Second, we examined the impact of different methods of calculating total costs for conditions attributable to diabetes. Our baseline AF estimates assigned equal shares of costs to diabetes and any of the attributable conditions listed for an event. To generate an “upper bound,” we assigned all costs to diabetes for events that listed diabetes, even if other attributable conditions were also listed. For example, if there was an MI event that listed both CVD and diabetes, our baseline split the costs evenly between CVD and diabetes, whereas the upper bound approach assigned all costs to diabetes. This algorithm, which clearly over-attributes costs to diabetes, led to an 11 percent higher AF estimate of \$41.3 billion.

Third, we examined the effect of using diabetes prevalence values from the National Health Interview Survey (NHIS). We used MEPS for baseline analyses to limit the possibility that differences in costs between RB and AF models were due to data differences. However, the NHIS is preferred for estimating diabetes prevalence (ADA 2003; CDC 2007). We used the NHIS to adjust the age- and condition-specific prevalence rates in equation (3) upward, which resulted in slightly higher AFs for each condition (details in Appendix S4). Estimated costs were \$39.2 billion—5.7 percent higher than baseline AF results.

Finally, when the previous two approaches were combined, both applying the “upper bound” AF algorithm and adjusting diabetes AFs upward with NHIS prevalence data, the estimated cost was 16.7 percent higher than our baseline AF estimate, or \$43.3 billion. This AF estimate, which appears to

over-attribute costs to diabetes, was virtually equal to our lowest RB estimate of \$43.3 billion. Because the upper bound AF estimate of \$43.3 billion is constructed to over-attribute costs to diabetes, while the lower bound RB estimate of \$43.3 billion likely underattributes costs to diabetes, we also examined the extent to which remaining cost differences are driven by higher treatment intensity for those with diabetes, which is captured in RB, but not AF, estimates.

As an example of this higher treatment intensity, we calculated the average length of hospital stay for CVD events by diabetes status. We found that, on average, people with diabetes spent 1.4 days longer in the hospital for CVD events than people without diabetes. This difference was statistically significant ($p = .0031$). For renal, endocrine/metabolic, PVD, and neurological disease events, people with diabetes had longer stays in the hospital than people without diabetes, although the differences were not statistically significant. These results suggest that the presence of diabetes complicates treatment for nondiabetes conditions and raises costs.

DISCUSSION

Applying the RB approach to estimate diabetes-attributable medical spending resulted in cost estimates that were 43 percent higher than estimates using an AF approach. The 95 percent confidence interval around our RB estimate (\$47–\$59 billion) does not overlap the 95 percent confidence interval of our AF estimate (\$34–\$40 billion), indicating that the two estimates are statistically different from each other (Schenker and Gentleman 2001). We explored several alternative explanations for the differences in cost estimates between the two approaches, but costs from the RB approach were consistently higher than those from the AF approach except when specifications were designed to over-attribute costs to AF models and underattribute costs to RB models. The higher RB estimates may be driven in part by our ability in the RB models to capture increased medical costs that result from higher treatment intensity among those with diabetes.

Our study had several limitations. First, although the RB approach is straightforward to implement, it requires the availability of data that include measures of demographic, clinical, and health system confounders. It further requires that researchers run several preliminary analyses to identify the most appropriate model for their data, which adds complexity, because there is no single best econometric specification for estimating health care spending.

Rather, features of the data to be used in the analyses should be explored (e.g., distribution of the error variance, skewness, kurtosis) before selecting the analytic model. However, recommended methods for applying the RB approach are available in the health economics literature, and several resources are available to help researchers implement these methods (Manning and Mullahy 2001; Buntin and Zaslavsky 2004; Deb, Manning, and Norton 2005).

Second, the AF approach is limited in that no standardized methodology for estimating costs for the attributable conditions exists. Attributable condition costs may be estimated using an accounting approach, an incremental approach, or some combination of approaches. The cost estimates for a particular disease may differ considerably, depending on the cost attribution approach used (Ward et al. 2000). Further, the AF approach may be complicated by requiring selection of the most appropriate epidemiologic formulas for the data (e.g., Flegal, Graubard, and Williamson 2004).

Third, we used MEPS data for both approaches. MEPS provides an ideal dataset for applying the RB approach because it contains detailed person-level information on medical spending, but it has several limitations that made implementing the AF approach difficult. First, disease information in MEPS is limited to self-reported conditions that were later assigned three-digit ICD-9 codes. Those codes do not provide enough detail to identify all of the conditions attributable to diabetes or whether an individual had Type 1 or 2 diabetes. Second, when multiple conditions were reported for the same event in MEPS, we were unable to distinguish primary from secondary conditions, which limited our use of an accounting approach to estimate condition costs. Finally, MEPS excludes important components of medical costs, such as nursing home care, that may account for a large share of disease-attributable costs for conditions that disproportionately affect older and institutionalized people.

It is also important to understand why our medical spending estimates using the RB (\$52.9 billion) and AF approaches (\$37.1 billion) are lower than the oft-cited estimate of \$91.8 billion in direct medical spending (\$99.7 billion in 2004 dollars) from the ADA (2003). The ADA study used an AF approach and included categories of medical expenditures—nursing home and hospice care—that were excluded from our analysis because they were not in MEPS. ADA researchers also used diabetes-specific utilization measures to estimate costs for each attributable condition. Further, their estimates of AFs for CVD, other diabetes complications, and general medical conditions were higher than ours, which appear to result from the authors' use of odds ratios (ORs), rather than RRs, to calculate AFs.¹

CONCLUSIONS

We compared results from two common methodologies for modeling disease costs and found that the RB approach resulted in higher diabetes cost estimates than the AF approach. Although cost differences may, in part, be driven by the challenges of implementing RB and AF approaches in a similar manner, we also found evidence that RB estimates are higher because they account for the impact of higher treatment intensity (e.g., longer LOS) among those with the disease. We recommend using the RB approach for estimating disease costs whenever individual-level data on disease prevalence and medical spending are available and when there is reason to believe that the presence of the disease or risk factor affects treatment costs for other conditions, as in the case of diabetes.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: This research was supported by Grant No. 1 P30 CD000138-01 from the Centers for Disease Control and Prevention to the RTI-UNC Center of Excellence in Health Promotion Economics. We thank Justin Trogon for valuable comments and Susan Murchie for expert editorial support.

Disclosures: None.

Disclaimers: None.

NOTE

1. ORs and RRs are similar when calculated for a low-prevalence condition but may differ considerably when calculated for high-prevalence conditions (e.g., CVD, general medical conditions; Zhang and Yu 1998; McNutt et al. 2003). We assessed the impact by re-estimating our AF models using ORs versus RRs and found diabetes-attributable costs of \$71.0 billion, similar to the analogous ADA estimate of \$84.0 billion (2004 dollars; to increase comparability, spending on nursing home and hospice care have been subtracted from the ADA estimate because these are not included in MEPS). The remaining differences in study results are likely due to differences in approaches. When we used ORs, we estimated diabetes-attributable expenditures of \$24.5 billion for general medical conditions versus the \$2.6 billion estimate shown in Table 3. These results suggest that the use of ORs rather than RRs likely overstates diabetes-attributable costs in MEPS.

REFERENCES

- Agency for Healthcare Research and Quality. 2000. *Methodology Report #11: Sample Design of the 1997 Medical Expenditure Panel Survey Household Component*. Rockville, MD: Agency for Healthcare Research and Quality.
- . 2006. "MEPS HC-036: 1996–2004 Pooled Estimation File." Rockville, MD: Agency for Healthcare Research and Quality [accessed on September 11, 2007]. Available at www.meps.ahrq.gov/mepsweb/data_stats/download_data/pufs/h36/h36u04doc.pdf
- American Diabetes Association (ADA). 2003. "Economic Costs of Diabetes in the US in 2002." *Diabetes Care* 26 (3): 917–32.
- Benichou, J. 2001. "A Review of Adjusted Estimators of Attributable Risk." *Statistical Methods in Medical Research* 10: 195–216.
- Buntin, M. B., and A. M. Zaslavsky. 2004. "Too Much Ado About Two-Part Models and Transformation? Comparing Methods of Modeling Medicare Expenditures." *Journal of Health Economics* 23: 525–42.
- Centers for Disease Control and Prevention. 2007. "Diabetes Data and Trends—Prevalence of Diabetes." Atlanta: Centers for Disease Control and Prevention [accessed on February 20, 2008]. Available at http://www.cdc.gov/diabetes/statistics/prevalence_national.htm
- Coller, M., G. R. Harrison, and M. M. McInnes. 2002. "Evaluating the Tobacco Settlement Damage Awards: Too Much or Not Enough?" *American Journal of Public Health* 92 (6): 984–9.
- Deb, P., W. Manning, and E. Norton. 2005. "Modeling Health Care Costs and Counts." Powerpoint Presentation at the 2005 International Health Economics Association Meetings, Barcelona, Spain [accessed on March 27, 2007]. Available at <https://harrisschool.uchicago.edu/faculty/articles/iHEAminicourse.pdf>
- Finkelstein, E. A., H. Chen, T. R. Miller, P. S. Corso, and J. A. Stevens. 2005. "A Comparison of the Case–Control and Case–Crossover Designs for Estimating Medical Costs of Nonfatal Fall-Related Injuries among Older Americans." *Medical Care* 43 (11): 1087–91.
- Finkelstein, E. A., I. C. Fiebelkorn, and G. Wang. 2003. "National Medical Expenditures Attributable to Overweight and Obesity: How Much and Who's Paying?" *Health Affairs* Web Exclusive: W3-219-226. [May 14, 2003].
- Fishman, P. A., Z. M. Khan, E. E. Thomson, and S. J. Curry. 2003. "Health Care Costs among Smokers, Former Smokers, and Never Smokers in an HMO." *Health Services Research* 38 (2): 733–49.
- Flegal, K. M., B. I. Graubard, and D. F. Williamson. 2004. "Methods of Calculating Deaths Attributable to Obesity." *American Journal of Epidemiology* 160: 331–8.
- Gross, D. J., L. Alecxih, M. J. Gibson, J. Corea, C. Caplan, and N. Brangan. 1999. "Out-of-Pocket Health Spending by Poor and Near-Poor Elderly Medicare Beneficiaries." *Health Services Research* 34 (1, part 2): 241–54.
- Heffler, S., S. Smith, S. Keehan, C. Borger, M. K. Clemens, and C. Truffer. 2005. "U.S. Health Spending Projections for 2004–2014: Shifts in Funding for Prescription

- Drugs and Lower Private Health Insurance Premium Growth Are Expected Over the Next Ten Years." *Health Affairs* W5: 74–85.
- Hodgson, T. A., and A. J. Cohen. 1999. "Medical Expenditures for Major Diseases." *Health Care Financing Review* 21 (2): 119–64.
- Honeycutt, A. A., S. D. Grosse, L. J. Dunlap, D. E. Schendel, H. Chen, E. Brann, and G. A. Homsy. 2003. "Economic Costs of Mental Retardation, Cerebral Palsy, Hearing Loss, and Vision Impairment." In *Research in Social Science Disability (Vol. 3: Using Survey Data to Study Disability: Results from the National Health Interview Survey on Disability)*, edited by B. M. Altman, S. N. Barnartt, G. Hendershot, and S. A. Larson, pp. 207–28. Oxford: Elsevier.
- Lee, D. W., J. W. Meyer, and J. Clouse. 2001. "Implications of Controlling for Comorbid Conditions in Cost-of-Illness Estimates: A Case Study of Osteoarthritis from a Managed Care System Perspective." *Value in Health* 4 (4): 329–34.
- Manning, W. G. 1998. "The Logged Dependent Variable, Heteroscedasticity, and the Retransformation Problem." *Journal of Health Economics* 17 (3): 283–95.
- Manning, W. G., and J. Mullahy. 2001. "Estimating Log Models: To Transform or Not to Transform?" *Journal of Health Economics* 21: 461–94.
- Manning, W. G., J. P. Newhouse, N. Duan, E. B. Keeler, A. Leibowitz, and M. S. Marquis. 1987. "Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment." *American Economic Review* 77 (3): 251–77.
- McNutt, L. A., C. Wu, X. Xue, and J. P. Hafner. 2003. "Estimating the Relative Risk in Cohort Studies and Clinical Trials of Common Outcomes." *American Journal of Epidemiology* 157: 940–3.
- Miller, V. P., C. Ernst, and F. Collin. 1999. "Smoking-Attributable Medical Care Costs in the USA." *Social Science and Medicine* 48: 375–91.
- Mullahy, J. 1998. "Much Ado About Two: Reconsidering Retransformation and the Two-Part Model in Health Econometrics." *Journal of Health Economics* 17: 247–81.
- Rice, D. P. 1994. "Cost-of-Illness Studies: Fact or Fiction? [Commentary]." *Lancet* 344 (8936): 1519–20.
- Rockhill, B., B. Newman, and C. Weinberg. 1998. "Use and Misuse of Population Attributable Fractions." *American Journal of Public Health* 88 (1): 15–9.
- Schenker, N., and J. F. Gentleman. 2001. "On Judging the Significance of Differences by Examining the Overlap between Confidence Intervals." *American Statistician* 55 (3): 182–6.
- Thacker, S. B., R. M. Ikeda, K. E. Giesecker, A. B. Mendelsohn, S. H. Saydah, C. W. Curry, and Y. W. Yuan. 2005. "The Evidence Base for Public Health Informing Policy at the Centers for Disease Control and Prevention." *American Journal of Preventive Medicine* 29 (3): 227–33.
- U.S. Department of Labor, Bureau of Labor Statistics. Consumer Price Index [accessed on January 17, 2007]. Available at <http://data.bls.gov/cgi-bin/surveymost?cu>

- Ward, M. M., H. S. Javitz, W. M. Smith, and A. Bakst. 2000. "A Comparison of Three Approaches for Attributing Hospitalization to Specific Diseases in Cost Analyses." *International Journal of Technology Assessment in Health Care* 16 (1): 125–36.
- Zhang, J., and K. F. Yu. 1998. "What's the Relative Risk? A Method of Correcting the Odds Ratio in Cohort Studies of Common Outcomes." *Journal of the American Medical Association* 280 (19): 1690–1.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Appendix S1. Definitions of Conditions.

Appendix S2. Algorithm to Estimate Aggregate Condition Costs Using the AF Approach.

Appendix S3. Results from Age-Specific Logit and GLM Models of Health Care Utilization and Spending.

Appendix S4. Diabetes Prevalence Adjustment.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.