VALUE IN HEALTH 17 (2014) 90–97

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

Comparative Effectiveness Research/HTA

Evaluating Direct Medical Expenditures Estimation Methods of Adults Using the Medical Expenditure Panel Survey: An Example Focusing on Head and Neck Cancer

Diarmuid Coughlan, MPharm, MSc[1,*], *Susan T. Yeh, MSc*[2], *Ciaran O'Neill, PhD*[1], *Kevin D. Frick, PhD*[2]

[1]*Economics of Cancer Research Group, Department of Economics, National University of Ireland, Galway, Ireland;* [2]*Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA*

ABSTRACT

Objective: To inform policymakers of the importance of evaluating various methods for estimating the direct medical expenditures for a low-incidence condition, head and neck cancer (HNC). **Methods:** Four methods of estimation have been identified: 1) summing all health care expenditures, 2) estimating disease-specific expenditures consistent with an attribution approach, 3) estimating disease-specific expenditures by matching, and 4) estimating disease-specific expenditures by using a regression-based approach. A literature review of studies (2005–2012) that used the Medical Expenditure Panel Survey (MEPS) was undertaken to establish the most popular expenditure estimation methods. These methods were then applied to a sample of 120 respondents with HNC, derived from pooled data (2003-2008). **Results:** The literature review shows that varying expenditure estimation methods have been used with MEPS but no study compared and contrasted all four methods. Our estimates are reflective of the national treated prevalence of HNC. The upper-bound estimate of annual direct medical expenditures of adult respondents with HNC between 2003 and 2008 was $3.18 billion (in 2008 dollars). Comparable estimates arising from methods focusing on disease-specific and incremental expenditures were all lower in magnitude. Attribution yielded annual expenditures of $1.41 billion, matching method of $1.56 billion, and regression method of $1.09 billion. **Conclusions:** This research demonstrates that variation exists across and within expenditure estimation methods applied to MEPS data. Despite concerns regarding aspects of reliability and consistency, reporting a combination of the four methods offers a degree of transparency and validity to estimating the likely range of annual direct medical expenditures of a condition.

Keywords: direct medical expenditures, econometrics, head and neck cancer, matching, Medical Expenditure Panel Survey.

Introduction

State and federal health policy makers often seek some estimate of the economic burden of a disease to inform decisions regarding resource allocation for prevention or treatment. With different budgetary responsibilities, the needs for and uses of such data will vary. Regardless, a cost-of-illness (COI) study is the main vehicle for arriving at such estimates [1]. These studies usually include a combination of health care and related resource use, productivity losses, and "intangible" burden related to quality of life [2]. The perspective and methodology used can greatly affect cost estimates [3] and varies between studies. The specification of what constituted "cost" is an important consideration—Does cost translate to "charges" from providers or "expenditures" reimbursed by payers? [4] In the absence of guidelines or well-accepted standards on the methods for COI studies, there is a clear need to inform policymakers and other researchers of the different approaches and the subsequent interpretation of results [4]. Indeed, a review of asthma cost studies in the United States shows that a 10-fold range in medical and nonmedical estimates has been reported [5]. Despite numerous limitations, COI studies remain popular and are often quoted in the mass media to highlight the magnitude of a particular problem.

This article focuses on analyzing the direct medical expenditures component of a COI estimate for a relatively low-incidence, but topical condition— head and neck cancer (HNC). Because a subset of HNCs is caused by the human papilloma virus (HPV), the economic burden of HNC is likely to contribute to the HPV vaccination debate [6]. Previous economic studies of HNC were derived from nonnationally representative sources—Surveillance, Epidemiology, and End Results–Medicare [7] and managed-care population [8].

The Medical Expenditure Panel Survey (MEPS), with a nationally representative respondent population, is commonly used for the purpose of generating a COI estimate [9]. Based on

* *Address correspondence to*: Diarmuid Coughlan, Department of Economics, St. Mary's Building, University Road, National University of Ireland, Galway, Ireland.

E-mail: diarmuidcoughlan@gmail.com.

1098-3015/$36.00 – see front matter

http://dx.doi.org/10.1016/j.jval.2013.10.004

recommendations set out in Clabaugh and Ward's review [4], the MEPS is an appealing data source for analysts intent on informing public policy. The MEPS can link information on individuals and households to their use of and expenses for health care. That the data is publicly available, components of care are often verified, and a standardized metric of cost is used make MEPS particularly useful [4]. A systematic review of COI studies suggests a typology to describe the direct medical expenditures of any disease: 1) the sum of all medical expenditures; 2) the sum of all disease-specific expenditures for a person with the disease; 3) the difference in total expenditures between a group of individuals with a disease and a matched sample of those with similar characteristics; and 4) the incremental expenditures associated with a disease estimated by using a regression-based approach that includes an indicator comparing individuals with and without the disease [1]. Staff members at the Agency for Healthcare Research and Quality (AHRQ), which conducts the MEPS, have previously discussed methodological issues related to estimating the COI of diabetes [10] and obesity [11]. Given the lack of strict COI guidelines when using the MEPS, we undertook a literature review of recent MEPS studies to instruct our estimation methodology.

Literature Review

This review of the MEPS literature pertains to studies that report health care expenditure estimates. The following search terms were used: In PUBMED: ("methods" [MeSH terms] OR "method" [text word] OR "economics" [Mesh] AND ["Medical expenditure panel survey" OR "MEPS"]) and EMBASE: ("cost analysis"/exp OR "cost analysis" AND "Medical expenditure panel survey"). Other databases searched were Econlit, Web of Science, and Tufts CEA Registry. Our inclusion criteria consisted of articles that reported an annual per-respondent direct medical expenditure for a specific disease/condition between 2005 and 2012. The information elicited from available articles included the following: 1) direct medical expenditure estimating method as a subject of "validity," 2) model specification/diagnostic tests as a subject of "reliability," and 3) comorbidity measure as a subject of "consistency."

Thirty-eight studies met our inclusion criteria [5,10,12–47]. The review highlighted considerable heterogeneity in the methods used to estimate the direct medical expenditures associated with a condition. A detailed systematic review of COI studies that use the MEPS is warranted. No MEPS health care expenditure study reported a range of estimates using all four COI methods. Only eight studies reported estimates using more than one of these methods [10,14,17,19,24,40,41,47]. Regression models were the most popular method (31 studies) of estimating the effect of a condition on health care expenditures [5,10,12–46]. Five studies reported condition-specific expenditures [14,18,39–41] (attribution approach), and two studies used disease-related events to identify patients [13,31]. Three studies used just a matching approach [25,34,35], and five studies reported the summation of all medical expenditures associated with a condition approach [17,20,21,23,24]. There was considerable methodological heterogeneity among the regression models. For just positive expenditures, the generalized linear model (GLM) log link and gamma distribution [5,12,14,17,18,32,36,40–46] (14 studies) was the most popular method followed by the logarithm of expenditures in an ordinary least squares regression (9 studies) [10,13,20,30,31,37,38,42,43]. Of the GLM studies, only eight made reference to model specification and diagnostic tests [5,14,17,32,33,36,44,46].

In total, 26 studies accounted for comorbidities or made some type of risk adjustment [5,10,13,14,16,18,20,22–46]. Such methods included accounting for specific medical conditions, creating a count of chronic diseases, or using the Charlson comorbidity index. It has also been argued that theoretically comorbidities should be equally prevalent in populations of people with and without certain stand-alone diseases [40].

In conclusion, this literature review highlights issues with the validity of the estimation methods used, the reliability of the models developed in the absence of specification tests, and the lack of consistency in accounting for comorbidity. Methodologies, however, are becoming more sophisticated—use of instrumental variables [46] and the combination of matching and regression [5] to derive an estimate are novel and likely to be replicated with future MEPS expenditure estimation studies.

Case Study: HNC and the MEPS

A detailed description of the survey can be found elsewhere [48] and on the MEPS Web site. Briefly, the MEPS collects data on expenditures related to medical events such as inpatient stays, outpatient, emergency room and ambulatory visits, and prescribed medicines. In addition to household interviews, the MEPS includes a medical provider component, a follow-back survey that collects expenditure data from a sample of medical providers used by survey participants and is considered to be more accurate than a household survey and given priority in expenditure estimation [49]. Information on specific medical conditions is obtained in the MEPS interview by asking respondents which "health problems" had "bothered" each household member during the observation period. Also, respondents report the reason for each medical event. This method identifies respondents with HNC, which results in an estimate of the annual "treated prevalence." This would be distinct from incidence (establishing phase-of-care expenditures) and prevalence (which includes long-term survivors expenditures) cost-of-care estimates. The Clinical Classification Software system, a tool for clustering the approximately 17,000 *International Classification of Diseases, Ninth Revision* condition codes into 285 mutually exclusive and homogeneous categories, was used to identify respondents with HNC (Clinical Classification Software = 11).

As the annual number of cases of HNC in MEPS is smaller than the 100 observations that the AHRQ suggests for making national estimates, 6 years of data (2003–2008) were pooled to generate an analytic sample [50]. In this case, the "pooled weight" is the yearly person weight divided by the number of years (i.e., 6). All expenditures were inflation adjusted to 2008 dollars by using the medical component of the Consumer Price Index. MEPS pooled data produce "average annual" estimates based on "person-years." This is because the same respondent can be observed in 2 years of consolidated year files. Total expenditures for a medical event are defined as the sum of direct payments made by all payers.

SAS software, version 9.3 (SAS Institute, Inc., Cary, NC), and Stata software, version 11.2 (StataCorp, College Station, Texas), were used for statistical analyses. The analyses incorporated MEPS person-level weights and variance adjustment weights (strata and primary sampling unit) that enable estimates to be nationally representative.

Method 1: Identify All Patients with Diagnosis and Sum Medical Expenditures

The objective is to identify respondents with the condition and sum their medical expenditures. We considered the "treated prevalence" as being those who have a diagnosis of HNC with any medical event. We consider this to be the middle ground between respondents who have reported a diagnosis of HNC

without necessarily any medical events and those with HNC-specific medical events, who have been referred to as "affected prevalence" in a study on cancer survivors [35]. This method is straightforward but provides an estimate of all expenditures with HNC rather than isolating the costs specifically due to the disease; this necessarily overestimates expenditures attributable to HNC.

Method 2: Sum Disease-Specific Medical Expenditures (Attribution)

This method restricts its attention to medical expenditures related to the disease of interest. In the MEPS, this is achieved by using the consolidated file, the condition file, condition-event link, and events files including inpatient, emergency room, outpatient, office-based, home health, and pharmacy. We therefore defined an HNC "case" as an adult with an HNC (Clinical Classification Software = 11)-specific medical expenditure as per AHRQ MEPS workshop notes. (For more details, please contact the corresponding author). This method may underestimate the direct medical expenditures because it fails to include "spillover costs" attributed to a specific condition. An example of a spillover cost in cancer care would be a medical visit for chemotherapy-induced nausea, which will be coded as nausea and therefore unknown to the analyst to be cancer-related [51]. Undoubtedly, other medical expenditures (e.g., mental health issues) are attributable to a respondent's cancer. Given that we did not find any reliable epidemiologic estimates in the literature, we essentially did not attribute any expenditure fraction of other conditions to HNC.

Method 3: Estimates Incremental Expenditures by a Matching Approach

Comparing the medical expenditures of those with and those without the disease and attributing the differences in medical expenditures to the disease is logical. The essential difficulty, however, is finding a control group that is reasonably comparable. The objective of matching is to create a control that is "balanced" with the case on the covariates. Matching variables should be related to the condition (e.g., HNC) and the outcome (e.g., total medical expenditure) [52]. There is little guidance/consensus on how to specifically implement matching analyses with total expenditure as the dependent variable. The general advice is to think carefully about the set of covariates to include in the matching procedure and err on the side of including more rather than fewer [52]. Because expenditures are payments, covariates that may be related to a respondent's ability to pay for treatment should be included in the matching algorithm. Inherently, this approach introduces a degree of subjectivity into the estimate.

The matching variables used in the reported analyses are age, sex, race, insurance status (proxy for ability to pay), number of priority medical conditions (proxy for comorbidity), and year of data collection. All variables were given equal weight, and we allowed a 7-year age gap between cases and controls to ensure that a full 1:1 match was achieved. The legitimacy of using survey population weights in making national estimates, however, is debatable, with arguments existing around whether weights are meaningful in the context of matched samples [25].

GMATCH [53], a nearest neighbor matching routine without replacement (controls allowed to be used as a match only once) in SAS, commonly known as "Greedy Matching," was used in this analysis. GMATCH goes through the cases one at a time and picks the best control match on the basis of defined characteristics.

Method 4: Estimate Incremental Expenditures Using a Regression-Based Approach

The Andersen [54] conceptual model of health services utilization is a popular organizing framework. An example of a MEPS study that uses this model looks at serious psychological distress [26]. The model suggests that an individual's utilization of health services is a function of predisposing (e.g., age, sex, race/ethnicity, and education), enabling (e.g., poverty category, insurance, and region), and need (e.g., self-reported health status, number of priority chronic conditions, and smoking status) factors. Health care expenditures pose particular challenges for econometric modeling. The distribution of strictly positive expenditures is typically skewed, kurtotic, and heteroskedastic [55]. A variety of models have been used to analyze expenditure/cost data, with many analysts now presenting more than one model in their reports [56,57]. We report the most popular MEPS regression model, GLM. This is a popular approach to modeling health care expenditures but not without its limitations [58].

The GLM framework requires a link function that relates the conditional mean to the covariates and a distribution, to specify the relationship between the variance and the mean [59]. The most popular specification of the GLM for health care expenditures has been the log-link (Equation 1) with a gamma distribution (variance proportional to the square of the mean; $\lambda = 2$ in Equation 2) [60].

$$E[y|x] = f(x'\beta) = \exp(x'\beta) \text{In}(E[y|x]) = x'\beta \quad (1)$$

$$y \sim Var(y|x) \approx (E[y|x])^{\lambda} \quad (2)$$

We conducted diagnostic tests by using the "glmdiag" program [61] that performs the recommended modified Park test for the GLM family and the Pearson correlation test (checks for systematic bias in fit on raw scale), the Pregibon link test (checks linearity of response on scale of estimation), and the modified Hosmer and Lemeshow test (checks for systematic bias in fit on raw scale) for the GLM link [62]. The incremental expenditures were calculated by using the method of "counterfactual regression predictions," giving an attributable expenditure figure based on the difference between the entire sample having HNC and no respondent having HNC [60]. This attributable expenditure estimate is then multiplied by the population weight to give the national estimate of direct medical expenditures.

Results

The annual number of respondents who self-reported to have HNC in the MEPS data set (12–30 respondents per year, pooled =120) is small relative to the number of respondents in the survey (~30,000 respondents per year, pooled = 191,407). The "treated prevalence" consists of respondents with a medical expense in a given year—113 person-year cases. Of these, two respondents were children. We considered only adults in our analytic sample (see Table 1). Using our interpretation, the "treated prevalence" refers to 111 "person-year" cases, and the "affected prevalence" (HNC–specific expenditures) refers to 105 "person-year" observations. Therefore, 6 observations that reported to have HNC had medical events that were not related to HNC.

Total Expenditures: Methods 1 and 2

The national direct medical expenditure by method 1 was $3.18 billion compared to $1.41 billion by method 2. The other reportable information using MEPS data is the nature of health care utilization and the type of payer. Figure 1A,B shows that ambulatory care (office-based and outpatient visits) is the most common medical event. The mean HNC-specific ambulatory visit expenditure is $395 (95% confidence interval [CI] $329–$461). We note that

Table 1 – Comparison of demographic characteristics in MEPS adult respondents (>18 y) with and without head and neck cancer (2003–2008).

	Head and neck cancer (N = 111)	General adult sample (N = 131,041)	P
Age (y), mean ± SE	63.21 ± 1.11	46.16 ± 0.05	<0.0001
Sex			<0.0001
Male	76 (68)	59,850 (46)	
Female	35 (32)	71,191 (54)	
Race			0.474
White	89 (80)	100,338 (77)	
Black	18 (16)	20,922 (16)	
Asian	3 (3)	6,261 (5)	
Other	1 (1)	3,520 (3)	
Ethnicity			<0.0001
Hispanic	6 (5)	31,019 (24)	
Non-Hispanic	105 (95)	100,022 (76)	
Education			0.665
No degree	29 (26)	31,062 (24)	
GED/high school diploma	56 (50)	63,609 (49)	
Bachelor degree or higher	26 (23)	35,608 (27)	
Employment			<0.0001
Not employed	67 (60)	43,966 (34)	
Employed	39 (35)	85,451 (65)	
Poverty			0.406
Poor	22 (20)	21,841 (17)	
Near poor	10 (9)	7,825 (6)	
Low income	16 (14)	21,108 (16)	
Middle income	26 (23)	38,541 (29)	
High income	37 (33)	41,726 (32)	
Metropolitan statistical areas (MSA)			<0.0001
Non-MSA	25 (23)	23,185 (18)	
MSA	81 (73)	106,502 (81)	
Region			<0.0001
West	17 (15)	33,987 (26)	
Northeast	13 (12)	19,788 (15)	
Midwest	22 (20)	25,794 (20)	
South	54 (49)	50,118 (38)	
Health insurance			<0.0001
Uninsured	1 (1)	24,905 (19)	
Public only	50 (45)	80,433 (61)	
Any private	60 (54)	25,703 (20)	
Self-reported health status			<0.0001
Poor/fair	49 (44)	20,334 (16)	
Good/very good/excellent	50 (45)	108,863 (83)	
Not ascertained	12 (11)	1,844 (1)	
No. of priority chronic conditions			<0.0001
0	30 (27)	71,469 (55)	
1–2	46 (41)	45,327 (35)	
3–5	25 (23)	12,872 (10)	
6+	10 (9)	1,373 (1)	
Currently smoking			0.867
No	82 (74)	94,110(72)	
Yes	20 (18)	24,685 (19)	
Not ascertained	9 (8)	12,246 (9)	
Sum of weights	215,662	219,364,212	

Note. Values are n (%) unless indicated otherwise.
GED, General Educational Development; MEPS, Medical Expenditure Panel Survey; SE, standard error.

inpatient events account for merely 2% of all events but for more than 40% of expenditures in this sample. The mean HNC-specific inpatient expenditure is $13,291 (95% CI $3,212–$23,369).

With MEPS, policymakers can get an understanding of the economic burden and plan accordingly. For HNC-related events, more than 40% is paid for by private health insurance and Medicare accounts for 35% of expenditures (Fig. 1C).

Incremental Expenditures: Matching Approach (Method 3)

Our definition of "treated prevalence" takes into account any medical event and not just disease-specific expenditures. Most MEPS analyses consider the incremental expenditure approach to be estimates based on respondents who reported the condition of interest. The literature review, however, also highlighted two

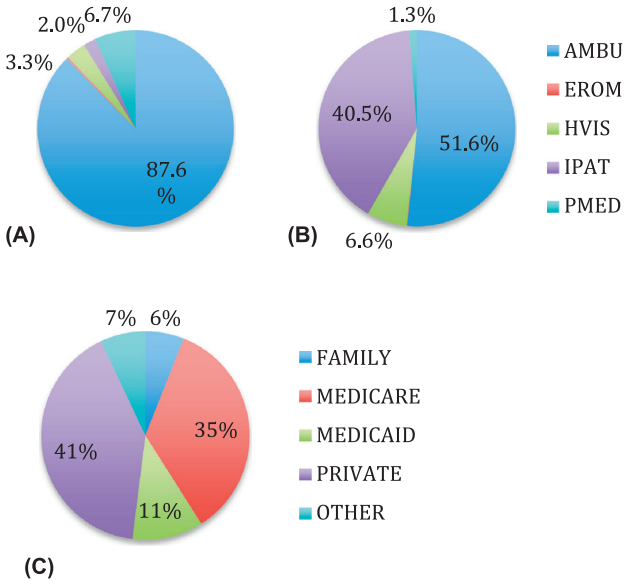

Fig. 1 – There were 978 events associated with head and neck cancer. (A) % by event type, (B) % by expenditure amount, and (C) the distribution of HNC events by payer. Code: AMBU, ambulatory visits (outpatient and office-based medical provider visits); EROM, emergency room visits; HVIS, home health visits; IPAT, hospital inpatient stay; PMED, prescription medicine.

analyses that used respondents with disease-specific expenditures as their analytic sample for incremental expenditures [13,31].

Table 2 shows the difference in expenditures between the groups attributed to the presence of HNC. Seven controls had $0 expenditures, and the largest annual expenditure of a control observation was $64,042. In comparison, the expenditure for cases ranged from $193 to $108,500.

Incremental Expenditures: Regression Approach (Method 4)

The reported GLM regression estimate (Table 3) is based on a complete case analysis where all cases had positive expenditures. The modified Park test indicated that the expenditure variable followed a Poisson distribution. (Variance is proportional to mean, $\lambda = 1$ in Equation 2.) The log-link function, however, yielded significant *P* values for the Pearson correlation test, the Pregibon link test, and the Hosmer and Lemeshow test. There is no single test that identifies the appropriate link, and we would ideally hope that all the three tests would be consistent in yielding nonsignificant *P* values [62]. The GLM "counterfactual regression predictions" method put the incremental cost attributable to HNC at $5069 (95% CI $4481–$5658) per person. Extrapolating to the national population by using the pooled weights in the MEPS for each respondent with HNC, the "average annual" total medical expenditures were calculated to be $1.09 billion.

Discussion

We report the results of four methods of estimating the direct medical expenditures for a condition using the MEPS nationally representative data set. According to our analyses, the average annual (from 2003 to 2008), direct medical expenditures for adults with HNC are in the range of $1.09 to $3.18 billion (in 2008 dollars). The National Cancer Institute (NCI) reports that HNC has an annual treatment cost of $3.64 billion (2010 US dollars) [63]. The NCI estimate is based on an incidence phase-of-care approach, and annualized costs were estimated from Surveillance, Epidemiology, and End Results–Medicare linkage data [64]. Although there are no gold standard data sources for estimating the prevalence costs of cancer care, the use of the MEPS for national estimates has serious limitations associated with small sample size [65]. The intention of this study, however, was to highlight the reliability of expenditure estimates with respect to methodological variation.

Method 1 gives an upper-bound estimate of $3.18 billion per year. Because this method does not identify the incremental expenditures attributable to HNC, it is not directly comparable to the other approaches, a fact that policymakers must bear in mind. As noted in a study on rheumatoid arthritis, the unadjusted means are likely to be a multiple of the incremental health care expenditures [19]. Methods 2 to 4 attempt to estimate the additional health care costs associated with a condition and are more directly comparable. How they do this though varies, as noted in the discussion of the methods, a fact that may affect the estimates produced and their subsequent interpretation. Method 2 captures only disease-specific events that have been recorded. We noted that coding of MEPS medical events might be associated with more than one condition and that without accounting for this duplicating of expenditures, spurious "double counting" can occur. Estimates of cancer-attributable spending based on the identification of cancer-related encounters, however, are subject to coding errors—for example, chemotherapy-induced nausea. For this reason, statistical models (methods 3 and 4) do not rely on diagnosis or procedure coding, and instead compare expenditures between individuals [66]. We argue that the importance of capturing the specific disease-linked events (method 2) is that it identifies the affected prevalence and the nature of the disease burden. In MEPS, ambulatory care is the most common type of medical event (857 of 978, ~87%) associated with HNC. Inpatient care, however, accounts for 40% of expenditures based on 20 events (~2%), and private health insurance (41%) shoulders the majority of the financial burden in HNC.

For a low-incidence condition such as HNC, it is reasonable to assume that all expenditures other than dental that are directly attributable to the cancer are captured in the MEPS. We would expect this method to set a lower bound, because it does not capture any of the "spillover" or indirect medical expenditures associated with the disease, which an incremental expenditure would capture. For a condition with many secondary effects such as diabetes, epidemiologic formulas based on attributable

Table 2 – Matching of head and neck cancer cases based on MEPS variables*.

	N	Mean	95% confidence interval	Mean difference in expenditure and national estimate
HNC cases	111	$14,733	$13,557–$15,909	$7,251
Controls	111	$7,482	$6,551–$8,413	$1.81 billion†

HNC, head and neck cancer; MEPS, Medical Expenditure Panel Survey.
* Age, sex, race, insurance status, number of priority medical conditions, and year of data collection using the GMATCH routine.
† Used survey weights from cases to generate a national estimate.

Table 3 – Summary of results from the four cost-of-illness methods.

	No. of observations	Per "person-year" mean* ± SE	Direct medical expenditure national estimate*
Total expenditures			
Method 1: Sum all medical costs	111	$14,733 ± $2,006	$3.18 billion
Method 2: Disease-specific medical costs	105	$6,884 ± $1,600	$1.41 billion
Incremental expenditures			
Method 3: Matching GMATCH 1:1	111	$7,251 ± $450	$1.56 billion
Method 4: Regression GLM – Log-link and Poisson distribution†	111	$5,069 ± $3,614	$1.09 billion

GLM, generalized linear model; MSA, Metropolitan statistical areas; SE, standard error.

* US 2008 dollars.

† Covariates included in regression analysis: age, sex, race, ethnicity, education, employment, poverty status, MSA, region, health insurance, self-reported health status, number of priority chronic conditions, and currently smoking.

fractions are needed to give a more accurate estimate of the annual expenditures [14]. In the case of asthma, the MEPS disease-specific estimate is 39% of the regression-based estimate of the incremental direct cost of asthma in 2006 [32].

The incremental expenditure approach requires subjective judgment regarding the specification of the model to be used, which may potentially lead to serious bias in both over- and underestimating expenditures. The likelihood that selection bias associated with undiagnosed HNC will materially affect results is very low. It is unlikely that those with undiagnosed HNC would have utilization patterns distinct from those of patients without HNC because early diagnosis of HNC is often missed because of the nonspecific symptoms or symptoms commonly associated with benign conditions [67]. The GMATCH matching routine is practical when variables are measured as discrete values. We chose to match all cases in this analysis and gave each matching variable equal weight. While this is acceptable, it is not the only and may not be the optimal approach to analyze these data. Further research could compare other algorithms and test the use of MEPS population survey weights as a matching variable. Although GMATCH is a freely available SAS program, more sophisticated routines exist in R and STATA that use caliper or kernel density matching with propensity scores. Additional research using the MEPS data set could compare the various techniques within matching to see what difference they make to the reliability of an expenditure estimate.

Specifications of regression models are often based on a conceptual framework such as the Andersen model of health services utilization. Categorical measures of self-reported health status have been shown to be good predictors of the subsequent use of medical care and are included in our analyses [68]. Indeed, the case for health status as the source of latent heterogeneity in health care use is strong [69]. As seen in Table 1, there are large differences among those with and without HNC in a number of variables including self-reported health status and the number of "quality priority conditions" (QPCs). Patients with cancer typically have more comorbidities than do those without [70,71]. Therefore, the mean annual expenditure of respondents with HNC ($14,733) is likely to consist of a number of other conditions and not just this cancer. In our analysis, we constructed a proxy for comorbidity by using a count of QPCs. We chose this method because it is the most straightforward to implement. Prescribing consistency when adjusting for comorbidity is difficult, because no single comorbidity measurement appears to be the best predictor of health care expenditure [72]. We argue that some measure is needed and that using the count of QPCs as a proxy for comorbidity can at least be applied across MEPS expenditure studies quite easily.

A lack of detail on the justification for regression-based models was noted in the literature review. In recent years, GLM is the preferred regression method among MEPS analysts. In our analysis, the modified Park test indicated that the GLM use a Poisson distribution. The literature review showed that only one other study used a Poisson distribution as dictated by model diagnostics [33]. The specification tests (Link test, Pearson test, and Hosmer and Lemeshow test), however, did detect problems with our log-link model, which suggests using a nonstandard link function instead. An estimator that attempts to relax the limitation of prespecifying a link or distributional family is the extended GLM approach termed extended estimating equations (EEE). [56]. We did expect that the mean GLM estimate ($5,069) based on 111 person-years would be higher than the disease-specific estimate ($6884) based on 105 person-years. This could indeed be due to a misspecification problem with the regression model.

The reliability of regression models is dependent on the model used. Recently, an MEPS study compared six econometric models and based their reported model on goodness-of-fit statistics—root mean square error, mean absolute error, and the scale-free Theil's statistic (15]. The conventional wisdom is that no single model is best for all cases, but comprehensive model checking is recommended [73]. With the MEPS, future regression-based analyses can include a plethora of models in STATA because code has been made freely available and textbooks become available to compare models in terms of a host of performance parameters such as mean absolution prediction error and the Copas test [74]. It is also possible that a "gold standard" approach of constructing a matched cohort and then performing a regression analysis could be used for the MEPS data [1]. This suggestion has recently been implemented for asthma [5].

Conclusions

We report a range of estimates of the direct medical expenditures of HNC based on four established approaches using the MEPS. Indeed, many additional estimates can be generated with matching and regression methods. As models become increasingly more sophisticated, however, the barriers for policymakers of accessing COI estimates and using them in an intelligent fashion to inform policy will increase. While it is accepted that policymakers are not the only audience for such studies, the value of a consensus on methods seems evident. In the absence of this, the caveat emptor will increasingly ring hollow and the value of such studies shall increasingly be called into question, especially in an emotive topic such as human papilloma virus vaccination of boys owing to the economic burden of treating subsequent cancers.

Source of financial support: Diarmuid Coughlan is funded as a research fellow of the Health Research Board in Ireland in conjunction with the National Cancer Institute.

REFERENCES

[1] Akobundu E, Ju J, Blatt L, Mullins CD. Cost-of-illness studies: a review of current methods. Pharmacoeconomics 2006;24:869–90.
[2] Larg A, Moss JR. Cost-of-illness studies: a guide to critical evaluation. Pharmacoeconomics 2011;29:653–71.
[3] Segel J. Cost-of-illness studies—a primer. Available from: http://academic.research.microsoft.com/Paper/5509981. [Accessed June 21, 2013].
[4] Clabaugh G, Ward MM. Cost-of-illness studies in the United States: a systematic review of methodologies used for direct cost. Value Health 2008;11:13–21.
[5] Rappaport H, Bonthapally V. The direct expenditures and indirect costs associated with treating asthma in the United States. J Aller Ther 2012;3:1–8.
[6] Jemal A, Simard EP, Dorell C, et al. Annual report to the nation on the status of cancer, 1975–2009, featuring the burden and trends in human papillomavirus (HPV)–associated cancers and HPV vaccination coverage levels. J Natl Cancer Inst 2013;105:175–201.
[7] Lang K, Menzin J, Earle CC, et al. The economic cost of squamous cell cancer of the head and neck: findings from linked SEER-Medicare data. Arch Otolaryngol Neck Surg 2004;130:1269–75.
[8] Amonkar MM, Chastek B, Samant N, Teitelbaum A. Economic burden of resected squamous cell carcinoma of the head and neck in a US managed-care population. J Med Econ 2011;14:421–32.
[9] Medical Expenditure Panel Survey background. Available from: http://meps.ahrq.gov/mepsweb/about_meps/survey_back.jsp]. [Accessed December 16, 2012].
[10] Olin G, Machlin S, Rhoades J. Estimating the cost of illness: the case of diabetes. 2008. Available from: http://gold.ahrq.gov. [Accessed December 16, 2012]. Report No. 08001.
[11] Rhoades J. Alternative techniques for examining the cost of obesity using the 2003 Medical Panel Expenditure Survey. 2006. Available from: http://apha.confex.com/apha/134am/techprogram/session_18296.htm. [Accessed April 20, 2012].
[12] Trogdon JG, Finkelstein EA, Feagan CW, Cohen JW. State- and payer-specific estimates of annual medical expenditures attributable to obesity. Obesity (Silver Spring) 2012;20:214–20.
[13] Balu S, Thomas J III. Incremental expenditure of treating hypertension in the United States. Am J Hypertens 2006;19:810–6.
[14] Honeycutt AA, Segel JE, Hoerger TJ, Finkelstein EA. Comparing cost of illness estimates from alternative approaches: an application to diabetes. Health Serv Res 2009;44:303–20.
[15] Basu R, Krueger PM, Lairson DR, Franzini L. Lifetime medical expenditures among hypertensive men and women in the United States. Womens Health Issues 2011;21:246–53.
[16] Sullivan PW, Ghushchyan V, Ben-Joseph RH. The effect of obesity and cardiometabolic risk factors on expenditures and productivity in the United States. Obesity (Silver Spring) 2008;16(9):2155–62.
[17] Monheit AC, Vistnes JP, Rogowski JA. Overweight in adolescents: implications for health expenditures. Econ Hum Biol 2009;7:55–63.
[18] Fu AZ, Qiu Y, Radican L, Wells BJ. Health care and productivity costs associated with diabetic patients with macrovascular comorbid conditions. Dia Care 2009;32:2187–92.
[19] Simons WR, Rosenblatt LC, Trivedi DN. The economic consequences of rheumatoid arthritis: analysis of Medical Expenditure Panel Survey 2004, 2005, and 2006 data. J Occup Environ Med 2012;54:48–55.
[20] Liem O, Harman J, Benninga M, et al. Health utilization and cost impact of childhood constipation in the United States. J Pediatr 2009;154:258–62.
[21] Liptak GS, Stuart T, Auinger P. Health care utilization and expenditures for children with autism: data from U.S. national samples. J Autism Dev Disord 2006;36:871–9.
[22] Kotlarz H, Gunnarsson CL, Fang H, Rizzo JA. Insurer and out-of-pocket costs of osteoarthritis in the US: evidence from national survey data. Arthritis Rheum 2009;60:3546–53.
[23] Cisternas MG, Murphy LB, Yelin EH, et al. Trends in medical care expenditures of US adults with arthritis and other rheumatic conditions 1997 to 2005. J Rheumatol 2009;36:2531–8.
[24] Yelin E, Murphy L, Cisternas MG, et al. Medical care expenditures and earnings losses among persons with arthritis and other rheumatic conditions in 2003, and comparisons with 1997. Arthritis Rheum 2007;56:1397–407.
[25] Lurie IZ, Manheim LM, Dunlop DD. Differences in medical care expenditures for adults with depression compared to adults with major chronic conditions. J Ment Health Policy Econ 2009;12:87–95.
[26] Dismuke CE, Egede LE. Association of serious psychological distress with health services expenditures and utilization in a national sample of US adults. Gen Hosp Psychiatry 2011;33:311–7.
[27] Bhattacharyya N. Incremental health care utilization and expenditures for chronic rhinosinusitis in the United States. Ann Otol Rhinol Laryngol 2011;120:423–7.
[28] Bhattacharyya N. Incremental healthcare utilization and expenditures for allergic rhinitis in the United States. Laryngoscope 2011;121:1830–3.
[29] Mitra S, Findley PA, Sambamoorthi U. Health care expenditures of living with a disability: total expenditures, out-of-pocket expenses, and burden, 1996 to 2004. Arch Phys Med Rehabil 2009;90:1532–40.
[30] Frick KD, Gower EW, Kempen JH, Wolff JL. Economic impact of visual impairment and blindness in the United States. Arch Ophthalmol 2007;125:544–50.
[31] Sullivan PW, Ghushchyan VH, Slejko JF, et al. The burden of adult asthma in the United States: evidence from the Medical Expenditure Panel Survey. J Allergy Clin Immunol 2011;127:363–9.e1–3.
[32] Barnett SBL, Nurmagambetov TA. Costs of asthma in the United States: 2002-2007. J Allergy Clin Immunol 2011;127:145–52.
[33] Kamble S, Bharmal M. Incremental direct expenditure of treating asthma in the United States. J Asthma 2009;46(1):73–80.
[34] Yabroff KR, Warren JL, Schrag D, et al. Comparison of approaches for estimating incidence costs of care for colorectal cancer patients. Med Care 2009;47(7, Suppl. 1):S56–63.
[35] Short PF, Moran JR, Punekar R. Medical expenditures of adult cancer survivors aged <65 years in the United States. Cancer 2011;117:2791–800.
[36] Tangka FK, Trogdon JG, Richardson LC, et al. Cancer treatment cost in the United States: has the burden shifted over time? Cancer 2010;116:3477–84.
[37] Wang J, Dong Z, Hong SH, Suda KJ. A comparison of direct medical costs across racial and ethnic groups among children with cancer. Curr Med Res Opin 2008;24:847–58.
[38] Yoon D, Frick KD, Carr DA, Austin JK. Economic impact of epilepsy in the United States. Epilepsia 2009;50:2186–91.
[39] Blanciforti LA. Economic burden of dermatitis in US workers. J Occup Environ Med 2010;52:1045–54.
[40] Martin BI, Turner JA, Mirza SK, et al. Trends in health care expenditures, utilization, and health status among US adults with spine problems, 1997-2006. Spine 2009;34:2077–84.
[41] Martin BI, Deyo RA, Mirza SK, et al. Expenditures and health status among adults with back and neck problems. JAMA 2008;299(6):656–64.
[42] Gunnarsson C, Chen J, Rizzo JA, et al. The direct healthcare insurer and out-of-pocket expenditures of psoriasis: evidence from a United States national survey. J Dermatol Treat 2012;23:240–54.
[43] Gunnarsson C, Chen J, Rizzo JA, et al. Direct health care insurer and out-of-pocket expenditures of inflammatory bowel disease: evidence from a US national survey. Dig Dis Sci 2012;57:3080–91.
[44] Gaskin DJ, Richard P. The economic costs of pain in the United States. J Pain 2012;13:715–24.
[45] Kawatkar AA, Jacobsen SJ, Levy GD, et al. Direct medical expenditure associated with rheumatoid arthritis in a nationally representative sample from the Medical Expenditure Panel Survey. Arthritis Care Res 2012;64:1649–56.
[46] Cawley J, Meyerhoefer C. The medical care costs of obesity: an instrumental variables approach. J Health Econ 2012;31:219–30.
[47] Miller JD, Foster T, Boulanger L, et al. Direct costs of COPD in the U.S.: an analysis of Medical Expenditure Panel Survey (MEPS) data. COPD 2005;2:311–8.
[48] Cohen SB. Design strategies and innovations in the Medical Expenditure Panel Survey. Med Care 2003;41(7, Suppl.):III5–12.
[49] Fleishman JA, Cohen JW. Using information on clinical conditions to predict high-cost patients. Health Serv Res 2010;45:532–52.
[50] Yu WW, Machlin S. Examination of skewed health expenditure data from the Medical Expenditure Panel Survey (MEPS). 2004. Available from: http://meps.ahrq.gov/mepsweb/data_files/publications/workingpapers/wp_04002.pdf. [Accessed October 16, 2012].
[51] Howard DH, Molinari N, Thorpe KE. National estimates of medical costs incurred by nonelderly cancer patients. Cancer 2004;100:883–91.
[52] Stuart EA. Matching methods for causal inference: a review and a look forward. Stat Sci Rev 2010;25:1–21.
[53] Kosanke J, Bergstralh E. GMATCH matching algoritm. 2004. Available from: http://mayoresearch.mayo.edu/mayo/research/biostat/upload/gmatch.sas. [Accessed December 12, 2012].
[54] Andersen RM. Revisiting the behavioral model and access to medical care: does it matter? J Health Soc Behav 1995;36:1–10.
[55] Blough DK, Madden CW, Hornbrook MC. Modeling risk using generalized linear models. J Health Econ 1999;18:153–71.
[56] Basu A, Rathouz PJ. Estimating marginal and incremental effects on health outcomes using flexible link and variance function models. Biostatistics 2005;6:93–109.
[57] Hill SC, Miller GE. Health expenditure estimation and functional form: applications of the generalized gamma and extended estimating equations models. Health Econ 2010;19:608–27.
[58] Basu A, Manning WG. Issues for the next generation of health care cost analyses. Med Care 2009;47(7, Suppl. 1.):S109–14.
[59] Manning WG, Mullahy J. Estimating log models: to transform or not to transform? J Health Econ 2001;20:461–94.

[60] Jones AM. Training workshop in health econometrics: Melbourne Institute of Applied Economic and Social Research: The University of Melbourne. Available from: http://www.melbourneinstitute.com/health/WorkshopApr202010.html. [Accessed December 18, 2012].
[61] Health Services Research Unit [University of Pennsylvania Health System]. Statistical analysis of costs—Stata programs. Available from: http://www.uphs.upenn.edu/dgimhsr/stat-cstanal.htm. [Accessed October 17, 2012].
[62] Glick H, Doshi J. Statistical considerations in health economic evaluations: 2011. Available from: http://www.uphs.upenn.edu/dgimhsr/ispor11.amsc.htm. [Accessed October 17, 2012].
[63] Cancer trends progress report – 2011/2012 update. Available from: http://progressreport.cancer.gov/index.asp. [Accessed January 17, 2013].
[64] Mariotto AB, Yabroff KR, Shao Y, et al. Projections of the cost of cancer care in the United States: 2010–2020. J Natl Cancer Inst 2011;103(2):117–28.
[65] Yabroff KR, Warren JL, Banthin J, et al. Comparison of approaches for estimating prevalence costs of care for cancer patients: what is the impact of data source? Med Care 2009;47(7, Suppl. 1):S64–9.
[66] Miller VP, Ernst C, Collin F. Smoking-attributable medical care costs in the USA. Soc Sci Med 1999;48:375–91.
[67] Mehanna H, Paleri V, West CML, Nutting C. Head and neck cancer–part 1: epidemiology, presentation, and prevention. BMJ 2010;341:c4684–46.
[68] Van Doorslaer E, Wagstaff A, van der Burg H, et al. Equity in the delivery of health care in Europe and the US. J Health Econ 2000;19:553–83.
[69] Deb P, Trivedi PK. The structure of demand for health care: latent class versus two-part models. J Health Econ 2002;21:601–25.
[70] Ogle KS, Swanson GM, Woods N, Azzouz F. Cancer and comorbidity. Cancer 2000;88:653–63.
[71] Smith AW, Reeve BB, Bellizzi KM, et al. Cancer, comorbidities, and health-related quality of life of older adults. Health Care Financ Rev 2008;29:41–56.
[72] Farley JF, Harley CR, Devine JW. A comparison of comorbidity measurements to predict healthcare expenditures. Am J Manag Care 2006;12:110–9.
[73] Deb P, Manning W, Norton E. Modeling health care costs and counts—American Society of Health Economists (ASHEcon) workshops. 2012. Available from: http://www.cce.umn.edu/American-Society-of-Health-Economists/Workshops/index.html. [Accessed December 28, 2012].
[74] Jones A, Rice N, Bago d'Uva T, Balia S. Applied Health Economics (2nd ed.). New York: Routledge, 2013.