

Attribution of Health Care Costs to Diseases: Does the Method Matter?

Allison B. Rosen
University of Massachusetts and NBER

Ana Aizcorbe
Virginia Tech and Bureau of Economic Analysis

Tina Highfill
Bureau of Economic Analysis

Michael E. Chernew
Harvard University and NBER

Eli Liebman
Duke University and Bureau of Economic Analysis

Kaushik Ghosh
NBER

David M. Cutler
Harvard University and NBER

We are grateful to Arlene Ash for comments, and to the National Institute on Aging (P01 AG031098 and R37AG047312) and the National Center for Research Resources (UL1 RR000161) for research support.

Abstract

Cost of illness (COI) studies focus on allocating health expenditures to a comprehensive set of diseases. A variety of techniques have been used to allocate spending to diseases. In this paper, we compare spending attributed to diseases using three approaches: one based on the principal diagnosis listed on each encounter's claim, a second based on all diagnoses listed on the encounter, and a third based on decomposing a person's total annual spending to their conditions. The study sample is large: 2.3 million commercially insured individuals under age 65. Results indicate significant differences in the allocations from the different approaches. The two claims-based encounter approaches allocate 78% of overall spending to diseases, while the person approach allocates 95% of spending to diseases. The large unallocated spending in the claims-based approach is due largely to lack of diagnosis codes for prescription medications. Spending was concentrated in a small number of conditions; the 10 most expensive diseases account for 40% of total spending with the person approach and about 18% of spending with the primary-diagnosis and all-diagnoses encounter approaches. Future research needs to pay careful attention to the choice of method in allocating spending to diseases, especially when research uses prescription medication claims data.

While health care cost growth in the United States has slowed in the past few years (Hartman, 2015), health costs are projected to grow faster than the economy over the next decade (Cutler and Sahni, 2013; Sisko et al., 2014; Keehan et al., 2015) and are one of the biggest fiscal challenges to the nation. As such, policymakers and analysts regularly try to better understand the value of this spending, so as to target cost containment efforts to curb excess – rather than essential – spending.

Unfortunately, there is often a mismatch between the data that are available and what policymakers need. Current National Health Expenditure Accounts measure medical spending at the level of the payers (Medicare, Medicaid, private insurance, etc.) and recipient of funds (hospital, physicians' office, pharmaceutical company, etc.). However, measuring the value of medical spending requires relating expenditures to the health outcomes they produce. This is most readily done at the disease level. For example, the value of spending more on physicians may be reflected in outcomes of hospitalization, or in hospitalizations avoided. This will only be picked up by looking at treatment for particular conditions. Thus, accurate cost of illness (COI) studies which allocate national health expenditures to a comprehensive set of diseases are an essential part of health policy.

Despite the importance of COI studies for health policy, no methodological standards for such studies exist and, to date, no side-by-side comparisons of estimates formed using different methods have been published. We address this gap in this paper.

Cost-of-illness studies come in two broad flavors. Most COI studies are *disease-based*, working from the bottom up to allocate costs to a single or limited number of diseases; absent constraints on collective spending, substantial double counting may – and often does – result

(Koopmanschap, 1998; Bloom, 2001; Rosen and Cutler, 2009). In contrast, *general* COI studies start with a population's total health care spending (often total health sector spending) and allocate some fraction of the sector's expenditures to each disease in a comprehensive mutually exclusive set (Rosen and Cutler, 2007, 2009). By constraining spending to national totals and applying consistent methods across diseases, *general* COI estimates are conceptually more meaningful for policy purposes and are, therefore, the focus of ongoing federal efforts to understand the diseases driving health care cost growth (Aizcorbe et al., 2008, 2011a, 2012a, 2012b, 2013; Bradley et al., 2010, 2013; Dunn et al., 2013a, 2013b, 2014a, 2015; National Research Council, 2005, 2008, 2010; Song et al., 2009). This paper focuses on the methods used to obtain these *general* COI estimates.

General COI studies date back to the 1960's (Scitovsky, 1964, 1967; Rice et al., 1967a, 1967b) and have increased in volume over time (see, for example, Cooper and Rice, 1976; Berk et al., 1978; Rice et al., 1985; Hoffman et al., 1996; Hodgson and Cohen, 1999; Druss et al., 2001, 2002; Thorpe et al., 2004a, 2004b, 2005, 2006, 2007, 2010, 2013; Roehrig et al., 2009, 2011; and Starr et al., 2014). As these general COI studies have proliferated, so have the methods used to generate their cost estimates.

Historically, most general COI studies have allocated claims to particular diseases at the *encounter-level*, assigning spending based on the diagnoses coded on each encounter's claim (Rosen and Cutler, 2009; National Research Council, 2010). The ease with which costs are attributed to diseases is a major advantage of this approach – it is essentially an accounting exercise. However, encounter-level costing is fairly limited in its capacity to handle comorbidities and downstream complications. If a person with diabetes and hypertension is

prescribed an ACE-Inhibitor (which can treat either condition), to which disease should the visit's costs and the medication cost be attributed? If this patient has a heart attack several years later, is the subsequent spending a result of the diabetes, the hypertension or the heart attack? Another disadvantage of encounter-level costing is that it cannot allocate spending for which there are no valid claims or diagnosis codes. How will the ACE-Inhibitor cost get allocated if the pharmacy claim has no diagnosis – and most pharmacy claims do not? Perhaps the biggest disadvantage of encounter-level cost-of-illness estimates is that they are not readily compared to health outcomes, which are measured at the *person-level*.

As such, interest has increased in using econometric models to recast cost-of-illness estimates at the *person-level*. This approach uses regression analysis to allocate an individual's total annual spending to their complete set of medical conditions (as indicated on their medical claims from that year). As such, person-level costing may produce more valid estimates in patients with multiple chronic diseases, as expenditures for comorbidities and complications are better captured. Person-level costing also allows spending for which there are no valid claims or diagnosis codes to be allocated. But, person-level analysis may be sensitive to choosing appropriate time windows in measuring disease prevalence (current year vs. previous year), and subject to bias if unobservables (e.g., socioeconomic status, or SES) are correlated with disease and spending.

However, these advantages come at the cost of added complexity. There is no single best econometric approach for modeling health care costs, leaving the analyst to test and decide between different model specifications. Further, the regression assumes that comorbidities have an independent effect on spending unless appropriate interaction terms are

included in the models. Identifying the appropriate groups of co-occurring diseases is an empirical issue which requires clinical expertise. Despite these limitations, person-level costing is quite appealing conceptually, as it allows for more meaningful comparisons between health care spending and health outcomes (such as mortality and quality of life), thereby providing the critical link between spending and health needed to more systematically measure value.

While both encounter- and person-level COI allocation methods are increasing in use, there have been no side-by-side comparisons of estimates from the different approaches to date. In this paper, we apply three different allocation methods – two encounter-level approaches common in the literature and a person-level approach – to allocate a population’s annual medical expenditures to a common comprehensive set of diseases; and to investigate the impact of method choice on the mix of spending across diseases and, for individual diseases, the treated prevalence, cost per case, and overall disease spending.

Our data are from the 2006 MarketScan commercial claims and encounters database. We have randomly selected 2.3 million individuals under the age of 65 with commercial insurance and prescription drug coverage in 2006. Using these data, we attribute annual spending to diseases using three different COI allocation approaches used in the literature: (1) the *primary-encounter* approach identifies all health care encounters and attributes spending to the principal diagnosis coded on the corresponding claim; (2) the *all-encounter* approach assigns each encounter’s spending to a combination of all (not just the principal) diagnoses coded on the corresponding claim; and, (3) the *person* approach identifies all of a person’s health conditions and, using regression analysis, allocates total spending to the diseases they experienced.

We compare outputs of the three approaches on several criteria, including the portion of spending allocated, the mix of spending across diseases; and, for individual diseases, treated disease prevalence, cost-per-case, and overall disease spending. For each approach, we explore in more detail the 10 conditions contributing the most to total spending.

The three approaches vary both in how much and how spending was allocated. The two encounter approaches allocate 77.7% of overall spending to diseases, while the person approach allocated 94.9% of spending to diseases. Further, the mix of spending across diseases differs substantially by method. Spending was concentrated in a small number of conditions; the 10 most expensive diseases accounted for 40.4% of total spending with the person approach and 18.1% and 18.3% of spending with the principal-diagnosis and all-diagnoses encounter approaches, respectively. These differences are sufficiently big that they warrant very careful attention to the choice of method in any cost allocation study.

This paper is structured as follows. Section 1 provides a review of the literature on different techniques used in measuring health care spending. In section 2, we discuss the different methodologies used in this study. In section 3, we explain our results. Section 4 discusses our findings and concludes.

I. Literature Review

In this section, we describe the methods that have been used to allocate total spending to diseases. We do so in parts.

Primary-encounter approach

The cost of illness studies dates back to sixties. A seminal study by Rice (1967) presented single-year estimates of health expenditures by type of disease for the year 1963. This study categorized diseases using International Classification of Diseases Adapted (ICDA). The total National Health expenditure in 1963 was estimated to be around \$22.5 billion. The diseases with highest spending were: the diseases of the digestive system (18.5%); mental, psychoneurotic and personality disorders (10.7%); and the diseases of the circulatory system (10.1%).

This study and the subsequent “cost of illness” literature in the 1960s, 1970s, and 1980s measured the total costs of illness in two dimensions: direct cost – which includes spending for different services including hospital, nursing home, physicians, medical professional services, drugs, medical supplies, research, training, and other non-personal – and indirect costs on morbidity and mortality, which account for economic losses arising from illness, disability, and death. Our focus in this paper is on direct costs.

Cooper et al. (1976) estimated that in 1972, the total cost of illness was \$188, out of which \$75 billion was direct cost, and for indirect cost - \$42 billion for morbidity and \$71 billion for mortality. Berk et al. (1978) estimated that the direct and indirect cost continued to increase, reaching \$264 billion dollars in 1975, with the diseases of digestive system, the diseases of circulatory system and mental disorders being the most expensive disease categories. Rice et al (1985) estimated the total economic cost of illness were \$455 billion in 1980. Other major studies in the 1970s and 1980s include Scitovsky (1985) and Hoffman et al. (1996).

But the biggest challenge in the 1960s and 1970s in measuring the cost of illness by disease was the lack of comprehensive and quality data on medical diagnoses and detailed spending breakdowns. Also, sophisticated econometric and statistical methods commonly used now to measure health care spending were not readily available. Most studies attempting to measure disease-based health care spending relied on the principal diagnosis on medical claims to assign spending to disease categories. These estimates were often overestimated or underestimated due to the presence of comorbid conditions. Starting in the mid to late 1990s, as more detailed data became available, researchers have been able to disaggregate spending more comprehensively.

One such study using the newer data sets in the late 1990s was by Hodgson and Cohen (1999). Hodgson and Cohen (1999) allocated 87 percent of personal health care expenditures as reported by the former Health Care Financing Administration (now CMS) by age, sex, diagnosis, and health service type using additional data from sources such as the National Medical Expenditure Survey. The diseases were classified using International Classification of Diseases, Ninth Revision (ICD-9) codes. Further disaggregation included home health care and hospital care by type of hospital. The diseases of the circulatory system (including, for example, heart disease and hypertension) were the most expensive condition, accounting for 17% of total personal health care expenditure. The diseases of the digestive system were the second most expensive conditions totaling 11%. The other major categories were injuries and poisoning, nervous system and sense organ diseases, and respiratory diseases. The top 6 categories contributed to 66% of Personal Health Care spending. **Table 1** gives a detailed review of the literature on studies that used a primary-encounter approach.

All-encounter approach

Starting early 2000s, there has been a trend in identifying the sources of changes in health care spending, focusing on medical conditions that make up a disproportionate amount of spending on health care and spending growth (for example, see Druss (2001, 2002), Thorpe et al. (2004a, 2004b), and Roehrig et al. (2009, 2011)). The studies by Thorpe and Rheorig were especially important as they looked at all diseases and their estimates were based on “all encounters” and not just the principal diagnosis coded on claims (i.e., “primary encounter”).

Thorpe et al. (2004a) used ICD-9 codes (truncated to 3 digits before inclusion in public-use national survey datasets) and subsequently coded them to 259 clinically relevant medical condition groupings using the Clinical Classification Software (CCS) developed by the U.S. Department of Health and Human Services (HHS). The authors started by pointing out that by using only the principal diagnosis, spending for some conditions will be understated. For example, diseases like hypertension, hyperlipidemia and diabetes will likely be underestimated using only the primary diagnosis as they are major comorbid conditions for acute events like heart attack, stroke, and renal failure.

To avoid such biases, Thorpe et al (2004a) proposed an estimation technique which has maximum (upper bound) and minimum (lower) bounds on cost estimates, and also proposed a novel estimation technique called “best guess”. Their upper-bound estimate attributed total spending to *each* health care event for which a given condition is listed. Since many medical conditions (up to 14) can be reported for each event, this will obviously include some double-counting. As a lower bound, they summed spending from each medical event for which only a single condition is reported. Although the total spending calculated from this approach

obviously does not account for all spending associated with a given condition, it does not include any double-counting.

Finally, they developed a “best guess” estimate of condition attributable spending using the following approach. They tabulated spending per event for those reporting a single medical condition. They then tabulated spending per event for those reporting two or more medical conditions associated with the event. They calculated the ratio of these two spending totals from single-diagnosis claims and used this to determine how much of the spending for claims with multiple conditions should be attributed to each individual condition.

Roehrig et al. (2009) in a similar and more comprehensive effort provided health expenditure estimates from the National Health Expenditure Accounts (NHEA) distributed across medical conditions. The study allocated spending to medical conditions using the nationally representative Medical Expenditure Panel Survey (MEPS) for the community population from 1996 to 2005. In addition, it provides guidance in identifying data and methods that cover the full range of expenditures in the National Health Expenditure Accounts (NHEA). Roehrig et al. found that the diseases of the circulatory system had the highest spending, accounting for 17% of total spending in 2005.

Roehrig et al. (2011) found that between 1996 and 2006, 75% of the increase in real per-capita health care spending was attributable to growth in cost per case, while treated disease prevalence accounted for 25% of spending growth. **Table 2** gives more detail on studies using an “all-encounter” approach to attribute health care costs to diseases.

Although, the “best-guess” approach addresses many of the concerns of the “primary encounter” method, it still has some limitations. First, it lacks a solid statistical or econometric

framework. Second, it is heavily dependent on finding claims with a single diagnosis for all medical conditions. At times, it is hard to satisfy this criterion for major claims like hospital visits and nursing home stays (which are often associated with multiple comorbid conditions). Finally, it is very difficult to assign prescription dollars to a medical condition, as prescription drugs claims do not include diagnosis codes. Next, we discuss a variant of encounter-based cost, referred to as an episode-based approach, which can address these issues and has been getting more popular in recent studies.

Episode-based approach

Increasingly, analysts are estimating disease costs using episode groupers—software programs with algorithms that organize claims from different sources (hospitals, nursing homes, physicians, hospital outpatient, home health, hospice, durable medical equipments and other medical services) for a given period of time (usually 6 months to a year) into distinct episodes of care that are clinically meaningful. Episodes are natural to examine because they group related claims regardless of where the service was provided; if a person is hospitalized for heart attack and stayed at a nursing home and then seen in followup at a physician office , all costs are included in the episode of heart attack care.

The most recent research at the Bureau of Economic Analysis (Dunn et al., 2013a, 2013b, 2014a, 2014b, 2015; Aizcorbe et al., 2008, 2011a, 2011b, 2012a, 2012b, 2013) uses this alternative method for measuring spending by disease. These so-called episode groupers use computer algorithms that sift through medical claims data and allocate spending to over 500 types of distinct disease episodes. There are a few groupers available in the market. One

popular grouper is Optum Symmetry Episode Treatment Group (ETG). It is an episode grouper for medical and pharmacy claims. It provides a condition classification methodology that combines related services into medically relevant and distinct units describing complete and severity-adjusted episodes of care and associated costs. **Table 3** gives a detailed account of the studies by the US Bureau of Economic Analysis (BEA) that assign spending to medical conditions using the ETG grouper.

Episode-based cost estimates have their own challenges. Identifying the start and end points of an episode of treatment is not straightforward, and it often takes many iterations to identify the optimum window. Comorbidities and their joint costs pose challenges as well, just as with the encounter approach. Other limitations include lack of clear guidelines on how to handle episodes related to the care of chronic diseases (should the episode be 1 year or 2 years?), handling complications of treatment and a few medical treatments that clearly don't fall under specific episode of care (screenings, etc.).

Finally, while a number of different commercial episode groupers are already widely in used, they have received little scientific evaluation to date (McGlynn, 2008), and the small but growing body of research by CMS and others points to real differences in the output of different vendors' groupers (MaCurdy et al., 2008, 2009; Rosen et al., 2012).

Person-based Approach

The final approach to cost estimation regresses a person's total annual health care spending on indicators for the set of medical conditions that person had during the calendar year. The results of this estimation can then be used to infer the cost of different conditions.

The most common estimation method is ordinary least square (OLS). The dependent variable in these regressions is total health care spending for each person. The independent variables usually are dummy variables indicating the presence (or absence) of various medical conditions. Other control variables generally include age, sex, gender, race, etc. The coefficients on disease dummy variables are the ones of interest. The regression coefficient on a disease dummy variable is the incremental additional cost of that condition, controlling for the other conditions the person has.

Because of the regression framework, a person-based approach is likely to produce more reliable estimates for patients with multiple chronic conditions, as it better accounts for spending related to comorbidities and complications. Further, prescription drug spending is naturally included, given that costs are not assigned to the specific condition on that claim.

That said, a regression specification may be sensitive to how comorbidities are entered. A standard linear regression may not be right since it imposes additivity of joint conditions. If having one condition increases (or decreases) the costs of another, an adjustment is needed to ensure that condition-specific spending does not sum to more (or less) than the total. Another empirical issue is what interaction terms to include. For the most part, clinical expertise is needed to identify the appropriate group(s) of co-occurring diseases, which may represent a limitation for policy purposes. **Table 4** reviews some of the literature that used such a regression approach. Importantly, as yet, no published studies have used a regression approach to allocate health care spending to a *comprehensive* set of conditions; rather, published studies focus on one or limited number of conditions of interest.

Estimation Techniques in Person-based Approach

Medical spending data has very specific characteristics that create challenges in efficiently estimating health care spending using the regression approach. A few common data issues are heteroscedasticity, heavy tails, and zero-spenders. Several studies have proposed more efficient estimation techniques to handle these data problems (Manning et al., 1998, 2001, 2005; Buntin and Zaslavsky, 2004; and Basu, 2009).

Manning et al. (1998) showed that the possibility of heteroscedasticity raises issues about the efficiency of the Ordinary Least Square estimates. In such cases, they recommended using Generalized Linear Squares estimators to obtain efficient estimates of the coefficients, and to further make accurate inference statistics for the standard error of such coefficients. Also, in case of log transformed or any other transformed dependent variable, the authors suggest that the researchers need to check if the error term is heteroscedastic across treatment groups or depends on some combination of independent variables. They also recommend that if the error terms is heteroscedastic, then the researchers should try to determine the form of the heteroscedasticity and use that information to obtain an unbiased estimate of the retransformation factor in order to estimate the overall expected level of spending to the independent variables (eg. medical condition dummies).

Manning and Mullahy (2001) examined how well the alternative estimators behave econometrically in terms of estimation bias and accuracy when the health spending data are skewed or have other common health expenditure data problems (zero-spenders, heteroscedasticity, heavy tails, etc.). They couldn't clearly identify any single alternative that best suits all conditions examined. They present a simple algorithm for choosing among the

alternative estimators. Selecting the right estimator is important for most accurate estimation. Their recommendation is to begin with both the raw-scale and log-scale residuals from one of the consistent Generalized Linear Model (GLM) estimators.

Manning, Basu and Mullahy (2005) found that there are two broad classes of models that can be commonly used to address the econometric problems caused by skewness in the health spending data. In the person level analysis, often times researchers encounter common data issues like zero-spenders, heteroscedasticity, heavy tails. The two common solutions proposed by the authors to deal with such data problem are: (1) transformation to deal with skewness (e.g., ordinary least square (OLS) on $\ln(\text{spending})$); and (2) different weighting approaches based on exponential conditional models (ECM) and generalized linear model (GLM) approaches. In this paper, they discuss these two classes of models using the three parameter generalized Gamma (GGM) distribution, which includes OLS with a normal error, OLS for the log-normal, the standard Gamma and exponential with log link, and the Weibull. The GGM also provides a potentially more robust alternative estimator to the standard alternatives.

Buntin and Zaslavsky (2004) compare the performance of eight alternative estimators, including OLS and GLM estimators and one- and two-part models, in predicting Medicare costs. They found that four of the alternatives produce very similar results in practice. They then suggest an efficient method for researchers to use when selecting estimators of health care costs. They recommended that researchers considering alternative models where the probability of use *per se* is not of interest would do well to start with the one-part GLM models.

Basu (2009) finds that zero spenders and skewed positive expenditure data can be best handled by one- part or two-part generalized linear model (GLM) with a gamma distribution

and a log link. In the two- part model, they use a logit model to predict the probability of having any medical spending and then use a GLM model with a gamma distribution and a log link to estimate the level of expenditures, given positive spending. **Table 5** gives a detailed review of the literature on studies addressing different techniques to estimate health care spending under a regression framework.

II. METHODS

In the United States, most people (54 percent) were covered by a health insurance plan related to employment for some or all of 2006 (State Health Facts Online, The Henry J. Kaiser Family Foundation). About 26 percent were covered by government health programs, including Medicare, Medicaid and other public programs. About 16% of the population was uninsured. **Figure 1** shows the population distribution by insurance coverage in 2006. For our analysis, we focus on the population covered under employer sponsored insurance.

Data and Study Sample

Study data were drawn from the 2006 MarketScan Commercial Claims and Encounters Database from Truven Health, which included enrollment and claims data for approximately 31 million individuals with employer-sponsored health insurance, provided largely by very large employers. MarketScan Commercial Claims and Encounters Database consist of employer and health plan-sourced data containing medical and drug data for several million individuals annually.

Enrollees include employees, their spouses, and dependents who are covered by the policy. Healthcare for these individuals is provided under a variety of fee-for-service (FFS), fully capitated, and partially capitated health plans, including preferred and exclusive provider organizations (PPOs and EPOs), point of service plans, indemnity plans, and health maintenance organizations (HMOs). Medical claims are linked to outpatient prescription drug claims and person-level enrollment information. **Figure 2** provides a schematic diagram of the Truven Health MarketScan claims data.

The enrollment files provide patient demographics, enrollment periods, types of coverage, and presence of medication coverage. The claims files provide inpatient, outpatient and prescription drug claims, and include dates and types of services, diagnosis (ICD-9-CM) codes, and costs of services. The maximum number of diagnoses recorded varies by claim type. Hospitalization claims include up to 15 diagnoses; outpatient claims up to two diagnoses; and prescription drug claims do not contain diagnosis codes. **Table 6** gives an account of the relevant variables in the MarketScan data.

We restricted our analysis by randomly selecting approximately 3 million individuals under the age of 65 with commercial insurance and prescription drug coverage in 2006. We excluded 0.58 million individuals with capitated insurance plans and dropped those with negative spending. The final analytic sample included 2.3 million individuals with 71.7 million claims totaling \$8.89 billion in annual spending (in 2006 U.S. dollars).

Classification of Diseases

Our goal was to use each method to allocate the samples' total health care spending in 2006 to a common set of mutually exclusive diseases. For our common core set of diseases, we used the 2012 version of AHRQ's Clinical Classification Software (CCS) (Elixhauser et al., 2012). The CCS software maps the approximately 14,000+ ICD-9-CM diagnosis codes into 283 mutually exclusive, clinically meaningful groups; the 283 single-level groups can then be aggregated up to 18 multi-level CCS chapters.

Methods for Allocation of Spending to Diseases

We allocated spending to the 283 CCS groups using three different approaches, as described in the previous section. Each approach is characterized by its methodological choices across three domains: the unit of observation (encounter versus person), the method of allocating costs to diseases (accounting versus econometric), and the handling of comorbidities (using all diagnoses versus principal diagnosis only).

Encounter-based Allocations

We examine two different encounter-based allocation approaches; both use basic accounting to allocate each medical claim's costs into the 283 CCS disease groups. Following the methodology of Rice (1967a), Cooper and Rice (1976), and Hodgson and Cohen (1999), our first approach (which we refer to as *Primary-Encounter*) assigns all of the spending on a single medical encounter to the principal diagnosis coded on its claim. While this approach is straightforward, it does not take into account the contribution of comorbidities to costs.

Our second approach follows more recent peer-reviewed literature (Thorpe, 2004a; Roehrig et al., 2009, 2011) allocating a portion of each encounter's spending to each (not just the principal) diagnosis coded on its claim. For claims with multiple diagnosis codes, the claims' spending is assigned to the coded diagnoses in proportion to the ratio of spending reported on claims with only one diagnosis (for more detail, see Appendix to Thorpe et al., 2004a). This approach, which we refer to as *All-Encounter*, attempts to better address the contribution of comorbidities to costs.

Person-based Allocation

To implement the *Person* approach, we regress each individual's total annual health care spending on indicators for the presence of diseases, as identified by diagnosis codes in the concurrent year's claims. In the simplest Ordinary Least Squares (OLS) specification, the coefficient on each condition represents the incremental additional spending for a person with that condition relative to someone without it. To deal with the right-skewed data, we used OLS regressions on log total expenditures; prior to log transformation, we added \$1 to each person's spending to ensure inclusion of individuals with no spending in 2006. Results were retransformed into their natural units using a smearing estimator (Duan, 1983), and \$1 was subtracted from each person's spending prior to final reporting. In the case of two conditions (d_1 and d_2), the regression is: $\ln(1+y) = \beta_0 + \beta_1 d_1 + \beta_2 d_2 + \epsilon$.

The log specification implicitly assumes that spending caused by any disease is multiplicative relative to spending without that disease. Because the underlying equation is non-linear, however, this approach will not lead to total spending matching population totals.

To address this issue, we followed a methodology described by Trogon et al. (2007, 2008), which estimates expenditures associated with co-occurring diseases and reallocates these expenditures to individual diseases. In this method, the estimated coefficients from the log regression are first used to separate out the portion of patients' spending that can be attributed to the conditions coded in their medical claims. The "attributable spending" for a patient is calculated as his observed spending less what his spending would have been if he had no conditions divided by observed spending:

$$AF_j = (E[y | d_j] - e[Y | d_j = 0]) / E[y | d_j]$$

The attributable spending for each individual is then allocated to conditions using shares calculated from the estimated coefficients. In the case of two conditions, the share of expenditures that are allocated to condition 1, for example, is:

$$S_1 = [\exp(\beta_1 - 1)] / \{ [\exp(\beta_1 - 1)] + [\exp(\beta_2 - 1)] \}$$

This method ensures that (i) all shares sum to one (i.e., all attributable spending is allocated), (ii) conditions with the larger coefficient are attributed a greater share of spending, and (iii) the only spending allocated to the patient are for conditions that the patient has.

Analyses

Analyses were restricted to the actual amounts paid for care for all claims completed during calendar year 2006. Charges are often reported on claims, but we do not use them. Because the encounter and the person allocation approaches are at different units of analysis – the individual claim and the person-year, respectively – we aggregated disease spending

estimates output by the two encounter approaches to the person-year to allow comparisons between the person and encounter estimates on a level playing field.

We started by comparing the proportion of total spending that each method was able to allocate to conditions. We then examined how each of the three methods distributed spending across CCS chapters. Then, for each CCS chapter, we examined differences in the number of patients with disease (*treated disease prevalence*), the average annual disease cost per patient with disease (*cost per case*), and the overall annual disease spending output by each allocation method. Finally, we examined in more detail the ten conditions accounting for the greatest share of total spending with each of the allocation method. All estimates are reported in 2006 dollars.

III. Results

Table 7 presents descriptive statistics for our study sample and their encounters (or claims). The study sample included 2.3 million commercially insured individuals with a mean age of 34; 51.3% are female. In 2006, the sample filed 71.7 million claims totaling \$8.89 billion in annual spending. This translated to a mean annual per person spending of \$3,788 (median \$1,640). The majority of claims (66.5%) were for outpatient services, with another 33.3 % for pharmacy services. Inpatient claims are a very small part of this sample. The average number of recorded diagnoses varied by claim type: 1 for outpatient services, 5 for inpatient services, and 0 for pharmacy claims.

The three methods differed in the portion of overall spending that could (and could not) be allocated to diseases, with far more spending allocated by the person method than by the

encounter methods (see first line in **Table 8**). Both encounter approaches had unallocated spending of \$1.98 billion (22.3 % of total). In contrast, the person approach had unallocated spending of \$450.0 million (5.1% of total). Over 99% of the unallocated encounter spending (\$1.97 billion) was for drug claims, which do not have diagnosis codes. In the person approach, unallocated spending is a result of unallocated constant.

The remaining rows of **Table 8** present, for each method, the treated disease prevalence, cost per case, and overall annual disease spending at the CCS chapter level. For all conditions, the treated disease prevalence is lower with the *primary-encounter* than with the *all-encounter* (or *person*) allocations. This is not surprising, as 10.7% of claims had more than one diagnosis coded. In contrast, the cost-per-case estimates from the two encounter methods were much closer than the estimates from the person approach. Diseases of the respiratory system provide an illustrative example: treated disease prevalence was 36.2% with the *person* and *all-encounter* allocations and 35.2% with the *primary-encounter* allocation; the cost-per-case was \$523, \$560 and \$956 from the *primary-encounter*, *all-encounter* and *person* allocation methods, respectively.

For any given disease, the overall disease spending estimated using the person approach often differed substantially from the estimates from either encounter approach, largely due to differences in the cost-per-case estimates. From our example above, total spending on diseases of the respiratory system was much higher with the person approach (\$813 million) than with the primary-encounter or the all-encounter approaches (\$432 and \$476 million, respectively). The mental health expenditures were higher with the person approach than with either encounter approach (\$333.3 million vs. \$225.7 and \$201.9 million)

perhaps indicating that comorbid conditions are better handled by regression approach. Total annual spending on neoplasms, on the other hand, was far higher with the primary- and all-encounter approaches (\$810 million and \$775 million, respectively) than with the person approach (\$434 million).

Figure 3 shows a radar plot of spending attributed to 18 Broad ICD9 disease categories by the all-encounter approach and the person-based approach. The biggest difference in attributable spending between the two methods is for “symptoms, signs, ill-defined conditions/factors influencing health”. There is a big un-attributable spending under claims based all-encounter approach.

Spending was concentrated in a small number of conditions. **Table 9** shows, for each allocation method, the 10 diseases (out of 283 CCS groups) accounting for the greatest total spending – and the spending on those conditions estimated by each of the other methods. The 10 most expensive diseases output by the person method accounted for 40.4% of total spending. In contrast, the 10 most expensive diseases output by the encounter methods accounted for 18.1% of spending when based on primary diagnosis alone and 18.3% of spending when all diagnoses were used.

The top 10 most expensive diseases differed by method (**Table 9** notes their ranking with each method). Both the primary- and all-encounter approaches identified ‘spondylosis, intervertebral disc and other back problems’ as the most expensive condition with overall spending of \$390 and \$358 million, respectively (versus \$304 million by person approach), and ‘coronary atherosclerosis and other heart disease’ as second most expensive with overall spending of \$197 and \$182 million, respectively (compared to \$124 million by the person

approach). In contrast, the person approach attributed the most spending to the medical examination/evaluation bucket, with overall spending of \$941 million (compared to \$108 million from both encounter approaches). Essential hypertension was the second most expensive disease from the person approach with overall spending of \$521 million; neither the primary- nor all-encounter approaches ranked hypertension among its ten most expensive conditions (overall hypertension spending of \$56 and \$80 million, respectively). Several of the other top 10 most expensive conditions with the person approach were not among the 10 most expensive from either of the encounter approaches, including lipid disorders, uncomplicated diabetes mellitus, other upper respiratory infections, and screening for conditions. In contrast, the two encounter approaches had non-specific chest pain, and breast cancer among their 10 most expensive diseases, while the person approach ranked them as the thirteenth, and sixteenth most expensive, respectively.

IV. Discussion

The need for fundamental change both in the financing and delivery of healthcare, and in the measurement of health sector productivity has stimulated interest by payers, policymakers and statistical agencies in allocating national spending across a comprehensive set of diseases (National Research Council, 2005, 2008, 2010; Rosen and Cutler, 2007, 2009; Aizcorbe et al., 2008, 2011a, 2012a, 2012b, 2013; Song et al., 2009; Bradley et al., 2010, 2013; Dunn et al., 2013a, 2013b, 2014a, 2015). However, there are no methodological gold standards guiding the performance of these COI studies. Applying three different COI methods to the

same data, we found that choice of method impacted both how much spending could be allocated to diseases and how that spending was allocated. The distribution of spending across diseases differed by method. In turn, for individual diseases, treated disease prevalence, cost per case, and overall disease spending varied depending on the method used. Results were close for some diseases but quite disparate for others.

Past studies comparing person-level and encounter-level cost of illness approaches demonstrate that *COI for a given disease* can vary widely depending on the choice of method (Lipscomb et al., 1998; Honeycutt et al., 2009; Ward et al., 2000; Akobundu et al., 2006; Yabroff et al., 2009); importantly, these studies have largely been restricted to individual diseases (i.e., they are effectively disease-specific COIs). However, as the policy import of general COI studies grows (National Research Council, 2005, 2008, 2010; Rosen and Cutler, 2007, 2009; Aizcorbe et al., 2008, 2011a, 2012a, 2012b, 2013; Song et al., 2009; Bradley et al., 2010, 2013; Dunn et al., 2013a, 2013b, 2014a, 2015), so does the critical need for studies comparing the different cost allocation methods employed specifically in this context.

While the research comparing different cost allocation methods in the context of general COI studies is in its infancy, a number of ongoing studies are underway. Several working papers report that the allocation of spending to diseases and, in turn, the price indexes that rely on these disease spending estimates, may be sensitive to method employed (for examples, see: Aizcorbe et al., 2011b; Rosen et al., 2012; Hall and Highfill, 2013; Dunn et al., 2014b). Indeed, in the recent release of the Bureau of Economic Analysis' new experimental Health Care Satellite Account, Dunn and colleagues (2015) comment on the importance of such comparisons moving forward (this first account employed a primary-encounter approach).

In the current study, we saw large differences both in the distribution of spending across diseases and in the within disease spending totals between the person-level and encounter-level methods. For example, mental health expenditures were much higher with the person approach than with either encounter approach, perhaps indicating that mental health is picking up the costs of common comorbid conditions. This would be consistent with literature demonstrating that depression raises the costs of treating a number of different chronic conditions (Welch, 2009). In contrast, spending on cancers was far higher with both encounter approaches than with the person approach, perhaps reflecting physician coding practices (diagnoses of cancer tend to get carried over from the initial claim to all subsequent claims).

The major advantage of the encounter approaches is the ease with which costs are attributed to diseases. Disadvantages include unclear handling of comorbidities, unallocated spending (i.e., claims without diagnoses) and inability to meaningfully link costs to health outcomes. The person approach is conceptually more appealing because it addresses the disadvantages of the encounter approach; most importantly, it allows for meaningful comparisons between health care spending and health outcomes. However, this comes with the price of additional complexity. There is no single best econometric approach for modeling health care costs, leaving the analyst to test and decide between a number of different model specifications. That said, there is a rich economics literature which can help guide the choice of model and its implementation (Manning et al. 1998, 2001, 2005; Buntin and Zaslavsky, 2004; Basu and Manning, 2009; Mullahy 2009).

Despite their apparent strengths and weaknesses, there are no standard metrics with which to compare encounter and person-level methods. Therefore, the best approach may

depend on the question on hand, data available, and the needs of the target audience, among other things. For example, if the goal is to compare costs and health effects within a given disease, as is done in cost-effectiveness analyses, a person-based approach may be best. In contrast, if price index construction is the goal, federal agencies may find an encounter-based approach more meaningful initially, until they are ready to make quality adjustments. In the long-term, more empirical work is needed on what approaches work best in which situations.

While our study has many strengths, it also has some limitations. While this study has demonstrated clear differences between the three COI allocation methods, it cannot provide definitive guidance on the choice of a 'best' or 'most appropriate' method for any given purpose. Rather, payers and policymakers must weigh the pros, cons and potentially conflicting information provided by each method, making value judgments as to which will best suit their needs. Second, while other COI allocation methods exist, we can only speak to those examined in the current study. One notable method – the use of episode groupers to allocate spending to diseases – is not used herein. Finally, our study compared the three methods at a point-in-time (i.e., cross-sectionally) and cannot be used to further inform efforts to understand the impact of method choice on price indices or other inherently longitudinal questions.

In summary, as the need to demonstrate the value of our health care spending increases, interest in allocating economy wide spending to a comprehensive set of diseases is likely to increase. This paper demonstrates that the choice of method may have very real implications for both how much and how that spending gets allocated. Additional empirical work developing these methodological tools and conceptual work exploring their ideal use will maximize their policy relevance and use.

REFERENCES

- Aizcorbe AM, Retus BA, Smith S. Toward a Health Care Satellite Account. *Survey of Current Business*. 2008;88:24-30.
- Aizcorbe A, Nestoriak N. Changing mix of medical care services: stylized facts and implications for price indexes. *J Health Econ*. 2011a;30(3):568–74.
- Aizcorbe AM, Bradley R, Greenaway-McGrevy R, Herauf B, Kane R, Liebman E, Pack S, Rozenal L. Alternative Price Indexes for Medical Care: Evidence from the MEPS Survey. *Bureau of Economic Analysis (BEA) Working Paper WP2011–01*. Washington, DC: BEA, 2011b.
- Aizcorbe A, Liebman E, Cutler DC, Rosen AB. Household Consumption Expenditures for Medical care: An Alternate Presentation. *Survey of Current Business*. 2012a;92(6):34-48.
- Aizcorbe A, Liebman E, Pack S, Cutler DM, Chernew ME, Rosen AB. Measuring health care costs of individuals with employer-sponsored health insurance in the U.S.: comparison of survey and claims data. *Statistical Journal of the International Association for Official Statistics*. 2012b;28(1):43-51.
- Aizcorbe A. Recent Research on Disease-Based Price Indexes: Where Do We Stand? *Survey of Current Business*. 2013;93:9–13.
- Akobundu E, Ju J, Blatt L, Mullins CD. Cost-of-illness studies: A review of current methods. *Pharmacoeconomics*. 2006;24:869-890.
- Basu A, Manning WG. Issues for the next generation of health care cost analyses. *Medical Care*. 2009;47(suppl 7):S109-S114.
- Berk A, Paringer L, Mushkin SJ. The Economic Cost of Illness: Fiscal 1975. *Medical Care*. 1978;16(9):785-790.
- Bloom BS, Bruno DJ, Maman DY, Jayadevappa. Usefulness of UC cost-of-illness studies in healthcare decision making. *Pharmacoeconomics*. 2001;19(2):207-213.
- Bradley R., Cardenas, E., Ginsburg, D.H., Rozenal, L., Velez, F., 2010. Producing disease-based price indexes. *Monthly Labor Review*. 2010;133:20–28.
- Bradley R. Feasible Methods to Estimate Disease-Based Price Indexes. *Journal of Health Economics*. 2013;32(3):504–514.
- Buntin MB, Zaslavsky AM. Too Much Ado About Two-Part Models and Transformation? Comparing Methods of Modeling Medicare Expenditures. *Journal of Health Economics*. 2004;23:525–42.

Cooper BS, Rice DP. The Economic Cost of Illness Revisited. *Social Security Bulletin*. 1976;39:21-36.

Cutler DM, Sahni RN. If Slow Rate Of Health Care Spending Growth Persists, Projections May Be Off By \$770 Billion. *Health Affairs*. 2013;32(5):841-850.

Druss BG, Marcus SC, Olfson M, Tanielian T, Elinson L, Pincus HA. Comparing the national economic burden of five chronic conditions. *Health Affairs*. 2001;20(6):233-41.

Druss BG, Marcus SC, Olfson M, Pincus HA. The most expensive medical conditions in America. *Health Affairs*. 2002;21(4):105–111.

Duan N. Smearing Estimate: A Nonparametric Retransformation Method. *Journal of the American Statistical Association*. 1983;78(383):605-610.

Dunn A, Liebman EB, Pack S, Shapiro A. Medical Care Price Indexes for Patients with Employer-Provided Insurance: Nationally Representative Estimates from MarketScan Data. *Health Services Research*. 2013a;48(3):1173–1190.

Dunn A, Shapiro AH, Liebman E. Geographic variation in commercial medical-care expenditures: A framework for decomposing price and utilization. *Journal of Health Economics*. 2013b;32: 1153-65.

Dunn A, Liebman B, Shapiro A. Developing a Framework for Decomposing Medical Care Expenditure Growth: Exploring Issues of Representativeness, in: D. W. Jorgenson, et al. (Eds.), *Measuring Economic Sustainability and Progress*. Chicago, IL: The University of Chicago Press for the National Bureau of Economic Research, 2014a.

Dunn A, Liebman EB, Rittmueller L, Shapiro A. Defining Disease Episodes and the Effects on the Components of Expenditure Growth. Bureau of Economic Analysis (BEA) Working Paper WP2014–4. Washington, DC: BEA, 2014b.

Dunn A, Rittmueller L, Whitmire B. Introducing the new BEA health care satellite account. *Survey of Current Business*. 2015;95:1-21.

Elixhauser A, Steiner C, Palmer L. Clinical Classifications Software (CCS), 2012. U.S. Agency for Healthcare Research and Quality. Available: <http://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp>.

Finkelstein EA, Fiebelkorn IC, Wang G. National medical spending attributable to overweight and obesity: how much, and who's paying? *Health Affairs*. 2003;W3:219-226.

Finkelstein EA, Chen H, Miller TR, Corso PS, Stevens JA. A comparison of the case–control and case–crossover designs for estimating medical costs of nonfatal fall-related injuries among older Americans. *Medical Care*. 2005;43(11):1087-91.

Hall AE, Highfill T. A Regression Based Medical Care Expenditure Index for Medicare Beneficiaries. Bureau of Economic Analysis (BEA) Working Paper WP2013–4. Washington, DC: BEA, 2013.

Hartman M, Martin AB, Lassman D, Catlin A, and the National Health Expenditure Accounts Team. National health spending in 2013: Growth slows, remains in step with the overall economy. *Health Affairs*. 2015;34(1):150-160.

Hodgson TA, Cohen AJ. Medical expenditures for major diseases. *Health Care Financing Review*. 1999;21(2):119–64.

Hoffman C, Rice D, Sung HY. Persons with chronic condition: their prevalence and costs. *JAMA*. 1996;276(18):1473-9.

Honeycutt AA, Segel JE, Hoerger TJ, Finkelstein EA. Comparing cost-of-illness estimates from alternative approaches: an application to diabetes. *Health Services Research*. 2009;44(1):303-20.

Koopmanschap MA. Cost-of-Illness studies, Useful for health policy? *PharmacoEconomics*. 1998;14(2):143-148.

Keehan SP, Cuckler GA, Sisko AM, Madison AJ, Smith SD, Stone DA, Poisal JA, Wolfe CJ, Lizonitz JM. National Health Expenditure Projections, 2014–24: Spending Growth Faster Than Recent Trends. *Health Affairs*. 2015;34(8):1407-17.

Lipscomb J, Ancukiewicz M, Parmigiani G, et al. Predicting the cost of illness: a comparison of alternative models applied to stroke. *Med Dec Making*. 1998;18:S39 –S56.

MaCurdy T, Kerwin J, Gibbs J, et al. Evaluating the functionality of the Symmetry ETG and Medstat MEG software in forming episodes of care using Medicare data. Burlingame, CA: Acumen, LLC; 2008. Available from: www.cms.gov/Reports/downloads/MaCurdy.pdf.

MaCurdy T, Kerwin J, Theobald N. Need for risk adjustment in adapting episode grouping software to Medicare data. *Health Care Financ Rev*. 2009;30(4):33-46.

Manning WG. The logged dependent variable, heteroscedasticity, and the retransformation problem. *Journal of Health Economics*, 1998;17(3):283–95.

Manning WG, Mullahy J. Estimating LogModels: to transform or not to transform? *Journal of Health Economics*. 2001;21:461–94.

Manning WG, Basu A, Mullahy J. Generalized modeling approaches to risk adjustment of skewed outcomes data. *Journal of Health Economics*. 2005;24:465-488.

Mullahy J. Econometric modeling of health care costs and expenditures. *Med Care*. 2009;47(suppl 7):S104-S108.

National Research Council. *Beyond the Market: Designing Nonmarket Accounts for the United States*. Washington, DC: The National Academies Press, 2005.

National Research Council. *Strategies for a BEA Satellite Health Care Account: Summary of a Workshop*. Washington, DC: The National Academies Press, 2008.

National Research Council. *Accounting for Health and Health Care: Approaches to Measuring the Sources and Costs of Their Improvement*. Washington, DC: The National Academies Press, 2010.

Rice DP. Estimating the cost of illness. *American Journal of Public Health*. 1967a;57(3):424–440.

Rice DP, Horowitz LA. Trends in medical care prices. *Social Security Bulletin*. 1967b;30:13-28.

Rice DP, Hodgson TA, Kopstein AN. The Economic Costs of Illness: A Replication and Update. *Health Care Financing Review*. 1985;6(1):61-80.

Roehrig C, Miller G, Lake C, Bryant J. National health spending by medical condition, 1996–2005. *Health Affairs*. 2009;28:w358–367.

Roehrig CS, Rousseau DM. The growth in cost per case explains far more of US health spending increases than rising disease prevalence. *Health Affairs*. 2011;30(9):1657-1663.

Rosen AB, Cutler DM. Measuring medical care productivity: a proposal for U.S. National Health Accounts. *Survey of Current Business*. 2007;87(6):54-58.

Rosen AB, Cutler DM. Challenges in building disease-based national health accounts. *Med Care*. 2009;47(suppl 7):S7-S13.

Rosen AB, Liebman E, Aizcorbe AM, Cutler DM. Comparing Commercial Systems for Characterizing Episodes of Care. Bureau of Economic Analysis (BEA) Working Paper WP2012–7. Washington, DC: BEA, 2012.

Scitovsky AA. (1964) An Index of the Cost of Medical Care - A Proposed New Approach, in: S. J. Axelrod (Ed.), *The Economics of Health and Medical Care*, University of Michigan, Bureau of Public Health Economics, Ann Arbor. pp. 128-142.

Scitovsky AA. Changes in the costs of treatment of selected illnesses, 1951-65. *American Economic Review*. 1967;57(5):1182-1195.

Scitovsky AA. Changes in the Costs of Treatment of Selected Illnesses, 1971-1981. *Medical Care*. 1985;23(12):1345-57.

Sisko AM, Keehan SP, Cuckler GA, Madison AJ, Smith SD, Wolfe CJ, Stone DA, Lizonitz JM, Poisal JA. National health expenditure projections, 2013-23: faster growth expected with expanded coverage and improving economy. *Health Affairs*. 2014;33(10):1841-50.

Song X, Marder WD, Houchens R, Conklin JE, Bradley R. (2009) Can A Disease-Based Price Index Improve the Estimation of the Medical Consumer Price Index?, in: W. E. Diewert, et al. (Eds.), *Price Index Concepts and Measurements*. The University of Chicago Press for the National Bureau of Economic Research, Chicago.

Starr M, Dominiak L, Aizcorbe A. Decomposing growth in spending finds annual cost of treatment contributed most to spending growth, 1980–2006. *Health Affairs*. 2014;33:823–831.

State Health Facts Online, 2006, The Henry J. Kaiser Family Foundation. Available at: www.statehealthfacts.org.

Sturm R. The Effects Of Obesity, Smoking, And Drinking On Medical Problems And Costs. *Health Affairs*. 2002;21(2):245-253.

Thorpe KE, Florence CS, Joski P. Which medical conditions account for the rise in health care spending? *Health Affairs*, 2004a; 23:w437–445.

Thorpe KE , Florence CS, Howard DH, and Joski P. The impact of obesity on rising medical spending. *Health Affairs*. 2004b;24:w480-486.

Thorpe KE, Florence CS, Howard DH, Joski P. The rising prevalence of treated disease: effects on private health insurance spending. *Health Affairs*. 2005; 24:w317–w325.

Thorpe KE, Howard DH. The rise in spending among Medicare beneficiaries: the role of chronic disease prevalence and changes in treatment intensity. *Health Affairs*. 2006;25:w378-88.

Thorpe KE, Howard DH, Galactionova K. Differences in disease prevalence as a source of the U.S.-European health care spending gap. *Health Affairs*. 2007;26:w678-w686.

Thorpe KE, Ogden LL, Galactionova K. Chronic conditions account for rise in Medicare spending from 1987 to 2006. *Health Affairs*. 2010;29(4):718-24.

Thorpe KE. Treated disease prevalence and spending per treated case drove most of the growth in health care spending in 1987–2009. *Health Affairs*. 2013;32(5):851–8.

Trogdon JG, Finkelstein EA, Nwaise IA, Tangka FK, Orenstein D. The economic burden of chronic cardiovascular disease for major insurers. *Health Promot Pract*. 2007;8(3):234-242.

Trogdon JG, Finkelstein EA, Hoerger TJ. Use of econometric models to estimate expenditure shares. *Health Services Research*. 2008;43(4):1442-1452.

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(suppl 7):S56-S63.

Ward MM, Javitz HS, Smith WF, and Bakst A. A Comparison of Three Approaches for Attributing Hospitalizations to Specific Diseases in Cost Analyses. *International Journal of Technology Assessment in Health Care*. 2000;16(1):125-136.

Welch CA, Czerwinski D, Ghimire B, Bertsimas D. Depression and costs of health care. *Psychosomatics*. 2009;50(4):392-401.

Table 1: Major Cost of Illness Studies: Primary Encounter (*principal diagnosis*) Method

Study	Study Period	Medical Conditions/Behavioral factors	Findings
Scitovsky (1967)	1951-1965	Otitis media in children, fracture of the forearm, acute cystitis, treated hypertension, pneumonia proved by x-ray, duodenal ulcer, coronary occlusion, maternity care, acute appendicitis, and cancer of the breast.	From 1951-52 to 1964-65, the costs of treatment of diseases covered by this study increased more than the US Bureau of Labor Statistics medical care price index.
Rice (1967a)	1963	Categorized according to the International Classification of Diseases, Adapted (ICDA).	Total National Health expenditure in 1963 was estimated to be \$22.5 billion. Out of that: neoplasm (5.7%), mental , psychoneurotic and personality disorders (10.7%), diseases of nervous and sense organs (6.3%), diseases of circulatory system (10.1%), diseases of respiratory system (7%), diseases of digestive system(18.5%) , diseases of bones and organs of movement (7.6%) and “all others” (28%). The diseases of digestive system was the biggest contributor.
Cooper et al. (1976)	1972	Classified by International Classification of Diseases, Ninth Revision (ICD-9) codes	The estimated total cost of illness in 1972 was \$188 billion, \$75 billion for direct costs. The top three were: the diseases of digestive system (\$11 billion), diseases of circulatory system (\$10.9 billion) and mental disorders (\$7 billion). Also, \$42 billion for morbidity and \$71 billion for mortality. The diseases of circulatory system was the most costly, representing about 20 percent of all costs of illness.
Berk et al. (1978)	1975	Classified by International Classification of Diseases, Ninth Revision (ICD-9) codes	Estimation of the direct and indirect costs of illness showed that the upward trend into total costs continued, reaching \$264 billion in 1975. The direct cost was \$118.5 billion, and indirect cost was \$145.8 billion- \$57.8 billion or morbidity and \$87.9 billion on mortality. In direct

			cost, the top 3 were the diseases of circulatory system (\$16 billion), diseases of digestive system (\$14 billion) and mental disorders (\$9.4 billion).
Rice et al. (1985)	1980	Classified by International Classification of Diseases, Ninth Revision (ICD-9) codes	In 1980, the estimated total economic cost of illness was \$455 billion: \$211 billion for direct costs, \$68 billion for morbidity, and \$176 billion for mortality. Diseases of the circulatory system and injuries and poisonings were the most expensive. There were variations in the diagnostic distributions among the three types of costs and by age and sex. In direct cost, the top three were the diseases of circulatory system (\$32.4 billion), diseases of digestive system (\$30.9 billion) and mental disorders (\$19.8 billion).
Scitovsky (1985)	1971 and 1981	Otitis media, Forearm fractures, Pneumonia, Duodenal ulcer, Complete physical examination, Appendicitis, Maternity care, Myocardial infarction, Breast cancer	The earlier study by the author, covering the periods 1951-1964 and 1964-1971, showed that cost increased due to change in relatively low-cost ancillary services, such as laboratory tests and x-rays ("littleticket" technologies). This study showed that in the period 1971-1981, the use of these technologies barely changed, but the use of a number of new and expensive technologies ("big-ticket" technologies) came into use, which raised health care costs significantly.
Hoffman et al. (1996)	1987	Classified by International Classification of Diseases, Ninth Revision (ICD-9) codes	The study estimated about 90 million Americans in 1987 were living with chronic conditions; 39 million of whom were living with more than 1 chronic condition. In the non-institutionalized population, over 45% had one or more chronic conditions. The direct health care costs account for 75 percent of the US health care expenditures. For people with chronic conditions total costs projected to 1990 amounted to \$659

billion-\$425 billion for direct health care costs and \$234 billion in indirect costs.

Hodgson et al. (1999)

1995

Classified by International Classification of Diseases, Ninth Revision (ICD-9) codes

This comprehensive study estimated that the diseases of the circulatory system were the most expensive category, costing \$127.8 billion and accounting for 17 percent of all Personal Health Care Expenditure (PHCE). Diseases of the digestive system cost \$86.7 billion, accounting for 11 percent of aggregate PHCE. The other six most costly disease categories in descending order were mental disorders (\$71.4 billion and 9 percent), injuries and poisonings (\$69.0 billion and 9 percent), nervous system and sense organ diseases (\$63.3 billion and 8 percent), and respiratory diseases (\$59.3 billion and 8 percent). Together, these six disease groups accounted for almost 66% of all PHCE. Neoplasm, including all cancers, represents only about 5 percent of total PHCE.

Table 2: Major Cost of Illness Studies: All-Encounter or Proportional Method

Study	Study Period	Medical Conditions/Behavioral factors	Findings
Druss et al. (2001)	1996	Mood disorders (depressive and manic depressive disorders), diabetes, heart disease, hypertension, and asthma	Direct per capita health costs for treatment of condition (mean per capita costs of health services that a person identified as resulting from the specific condition) was mood disorder (\$1,122), diabetes (\$1,097), heart disease (\$6,463), hypertension (\$569) and asthma (\$663). Mean per capita health costs for persons with condition (all costs borne by persons with the particular condition, including both direct costs and costs for comorbid conditions) : mood disorder(\$4,328), diabetes(\$5,646), heart disease (\$10,823), hypertension(\$4,073) and asthma(\$2,779). Estimated total health costs (billions) for persons with condition: mood disorder (\$54.9), diabetes (\$54.2), heart disease (\$38.5), hypertension (\$110.3) and asthma (\$27.7).
Druss et al. (2002)	1996	Classified diseases based on slightly modified Global Burden of Disease categories	Spending for the fifteen highest-cost conditions accounted for 44.2 percent of total U.S. health care spending in 1996. The top 15 conditions in billions of dollars were : ischemic heart disease (\$21.5), motor vehicle accidents (\$21.2), acute respiratory infection(\$17.9), arthropathies (\$15.9), hypertension(\$14.8), back problems(\$12.2), mood disorders(\$10.2), diabetes (\$10.1), cerebrovascular disease (\$8.3), cardiac dysrhythmias (\$7.2), peripheral vascular disorders (\$6.8), COPD(\$6.4), asthma (\$5.7), congestive heart failure (\$5.2) and respiratory malignancies(\$5.0).

Thorpe et al. (2004a)	1987 and 2000	The ICD-9 codes are collapsed to three-digit codes and subsequently coded into 259 clinically relevant medical conditions using the Clinical Classification System (CCS) developed by the U.S. Department of Health and Human Services (HHS).	Estimates included upper bound, lower bound and best guess estimates .The top 15 conditions accounted for 56 percent spending growth, with a lower bound of 43% and upper bound of 61%. The top 15 conditions in descending order were heart disease (8.06%), pulmonary disease (5.63%), mental disorders (7.40%), cancer (5.36%), hypertension (4.24%), trauma(4.64%), cerebrovascular disease (3.52%), arthritis (3.27%), diabetes (2.37%), back problems(2.99%), skin disorders (2.26%), pneumonia (2.26%), infectious disease (1.35%), endocrine disease (1.18%) and kidney disease (1.03%).
Thorpe et al. (2004b)	1987 - 2001	The ICD-9 codes were collapsed to three-digit codes and subsequently coded into 259 clinically relevant medical conditions using the Clinical Classification System (CCS) developed by the U.S. Department of Health and Human Services (HHS).	Obesity attributable health care spending increased between 1987 and 2001. Increases in obesity prevalence alone account for about one-tenth of the growth in health spending. The study estimated that the increases in the share of and spending on obese individuals relative to individuals of normal weight account for one third of the rise in inflation-adjusted per capita spending between 1987 and 2001. Out of that: spending for diabetes, 38%; spending for hyperlipidemia, 22%; and spending for heart disease, 41%.
Roehrig et al. (2009)	1996– 2005	ICD-9 codes mapped into CCS categories. Additional categories for prevention/exams (general checkups, well-child visits, immunizations, eye exams, and disease-specific screening procedures) and dental care were added.	This study provided health expenditures from the National Health Expenditure Accounts (NHEA) distributed across medical conditions. It provided annual estimates from 1996 to 2005 for about 30 or so medical conditions combined into 13 all-inclusive diagnostic categories. Circulatory system spending was highest, accounting for 17 percent of spending in 2005. The most costly conditions were mental disorders and heart conditions. Spending growth rates

			were lowest for lung cancer, chronic obstructive pulmonary disease, pneumonia, coronary heart disease, and stroke. This slow growth in these diseases was attributed to benefits of preventive care.
Roehrig et al. (2011)	1996 and 2006	The distribution of spending by condition was made using the Clinical Classification System software—developed by the Agency for Healthcare Research and Quality(AHRQ)—which maps detailed diseases onto an all-inclusive set of 260 medical conditions	The authors examined treated prevalence, clinical prevalence—the number of people with a given disease, treated or not— and cost per case across all medical conditions between 1996 and 2006. Over this period, 75% of the increase in real per capita health spending was attributable to growth in cost per case, while treated prevalence accounted for about 25% of spending growth.

Table 3: Major Cost of Illness Studies: Episode Grouper Method

Study	Study Period	Medical Conditions/Behavioral factors	Findings
Dunn et al. (2013a)	2003 to 2007	Classified disease spending using a commercial algorithm called a grouper; specifically, the authors use the ETG grouper from Symmetry. The ETG grouper allocates each record into one of over 500 disease groups called “episode treatment groups” (ETGs).	Service Price Index (SPI) grew 0.7 percentage points faster than the preferred MCE (Medical Care Expenditure) index.
Aizcorbe et al. (2012b)	2005	All conditions	Both total spending and the distribution of annual per person spending differed across the two data sources, with MEPS estimates 10 percent lower on average than estimates from MarketScan. These differences appeared to be a function of both.
Dunn et al. (2014b)	2003 to 2007	Classified disease spending using a commercial algorithm called a grouper; specifically, the authors use the ETG grouper from Symmetry. The ETG grouper allocates each record into one of over 500 disease groups called “episode treatment groups” (ETGs).	The goal of this paper was to better to obtain nationally representative estimates of the various components of expenditure growth. Using a multitude of weighting strategies, including weighted and un-weighted estimates, the authors found similar qualitative results with higher prevalence and increases in medical care service prices being the key drivers of spending growth.

Dunn et al.
(2015)

2015

In this study, the MEPS account was constructed using data from the MEPS. Each encounter in the data includes expenditure information and a primary ICD-9 diagnosis code. Each diagnosis code was mapped into one of 263 possible CCS categories. In market Scan data, the authors apply a person-based approach to allocate expenditures across CCS disease categories (Dunn et al, 2014)

The main focus of this study was creation of "The Blended Account" to comprehensive account spending by medical conditions. Blended Account was to substitute pieces of the Medical Expenditure Panel Survey for certain populations (with inadequate or no data) with corresponding big data. The two data sets that they incorporate into the blended account are the Medicare and MarketScan data. The results show significant improvement in measurements adding this big data.

Table 4: Major Cost of Illness Studies: Person-based Allocations

Study	Study Period	Medical Conditions/Behavioral factors	Method	Findings
Sturm (2002)	1997–1998	Smoking, Drinking, Obesity	Regression Approach	Regression analysis showed that obese adults incurred annual medical expenditures that were \$395 (36 percent) higher than those of normal weight incur.
Finkelstein et al. (2003)	1996–1998	Overweight and Obesity	Regression Approach	Used regression approach and national data in 1998 to calculate aggregate overweight- and obesity-attributable medical spending for the United States and by select payers. Expenditures for this group accounted for 9.1 percent of total annual medical expenditures. Medicare and Medicaid paid about 50% of these costs.
Finkelstein et al. (2005)	1998 and 1999	Fall-Related Injuries	The case–control design using regression and case–crossover approach.	On average, the estimates of the costs of fall injuries from the case–control design were between 6% and 17% greater than those from the case–crossover approach.
Trogdon et al. (2008)	2000–2003	Other MH/SA, hypertension, diabetes, arthritis, dyslipidemia, heart disease, asthma, kkin disorders, depression and HIV	Per Person Expenditures (generalized linear model), Attributable Fraction (%) generalized linear model	The authors stated that “Incremental effects of conditions on expenditures, expressed as a fraction of total expenditures, cannot generally be interpreted as shares. When the presence of one condition increases treatment costs for another condition,

summing condition-specific shares leads to double-counting of expenditures. Condition-specific shares generated from multiplicative models should not be summed". The authors provide an algorithm that allows estimates based on these models to be interpreted as shares and summed across conditions.

Honeycutt (2009)	1998 to 2003	Diabetes	Regression based Approach , Attributable Fraction Approach	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).
------------------	--------------	----------	--	---

Table 5: Major Cost of Illness Studies: Person-based Allocations, Methodological Issues

Study	Study Period	Method	Findings
Manning et al. (1998)	1998	OLS and GLS	Manning et al. (1998) showed that the possibility of heteroscedasticity could raise major issues about the efficiency of the Ordinary Least Square estimates. In such cases, they recommended using Generalized Linear Squares estimators to obtain efficient estimates of the coefficients, and to further make accurate inference statistics for the standard error of such coefficients. Also, in case of log transformed or any other transformed dependent variable, the authors suggested that the researchers need to check if the error term is heteroscedastic across treatment groups or depends on some combination of independent variables. They also recommend that if the error terms is heteroscedastic, then the researchers should try to determine the form of the heteroscedasticity and use that information to obtain an unbiased estimate of the retransformation factor in order to estimate the overall expected level of spending to the independent variables (eg. Medical condition dummies).
Manning and Mullahy (2001)	2001	OLS, GLM	Manning and Mullahy (2001) examined how well the alternative estimators behave econometrically in terms of estimation bias and accuracy when the health spending data are skewed or have other most common data problems (zero-spenders, heteroscedasticity, heavy tails, etc.). They

couldn't clearly identify any single alternative that best suits all conditions examined. Although, they present a simple algorithm for choosing among the alternative estimators. Selecting the right estimator is important for most accurate estimation. Their recommendation is to begin with both the raw-scale and log-scale residuals from one of the consistent Generalized Linear Model (GLM) estimators.

Manning Basu , Mullahy (2005)	2005	Ordinary least square (OLS),exponential conditional models (ECM), generalized linear model (GLM)	Manning, Basu and Mullahy (2005) found that there are two broad classes of models that can be commonly used to address the econometric problems caused by skewness in the health spending data. In the person level analysis, often times researchers encounter common data issues like zero-spenders, heteroscedasticity, heavy tails. The two common solutions proposed by the authors to deal with such data problem are: (1) transformation to deal with skewness (e.g., ordinary least square (OLS) on ln(spending)); and (2) different weighting approaches based on exponential conditional models (ECM) and generalized linear model (GLM) approaches. In this paper, they also discussed these two classes of models using the three parameter generalized Gamma (GGM) distribution, which includes OLS with a normal error, OLS for the log-normal, the standard Gamma and exponential with a log link, and the Weibull. The GGM also provides a potentially more robust alternative estimator to the standard alternatives.
--	------	--	--

Buntin and Zaslavsky (2004)		OLS , GLM	Buntin and Zaslavsky (2004) compare the performance of eight alternative estimators, including OLS and GLM estimators and one- and two-part models, in predicting Medicare costs. They find that
--------------------------------------	--	-----------	--

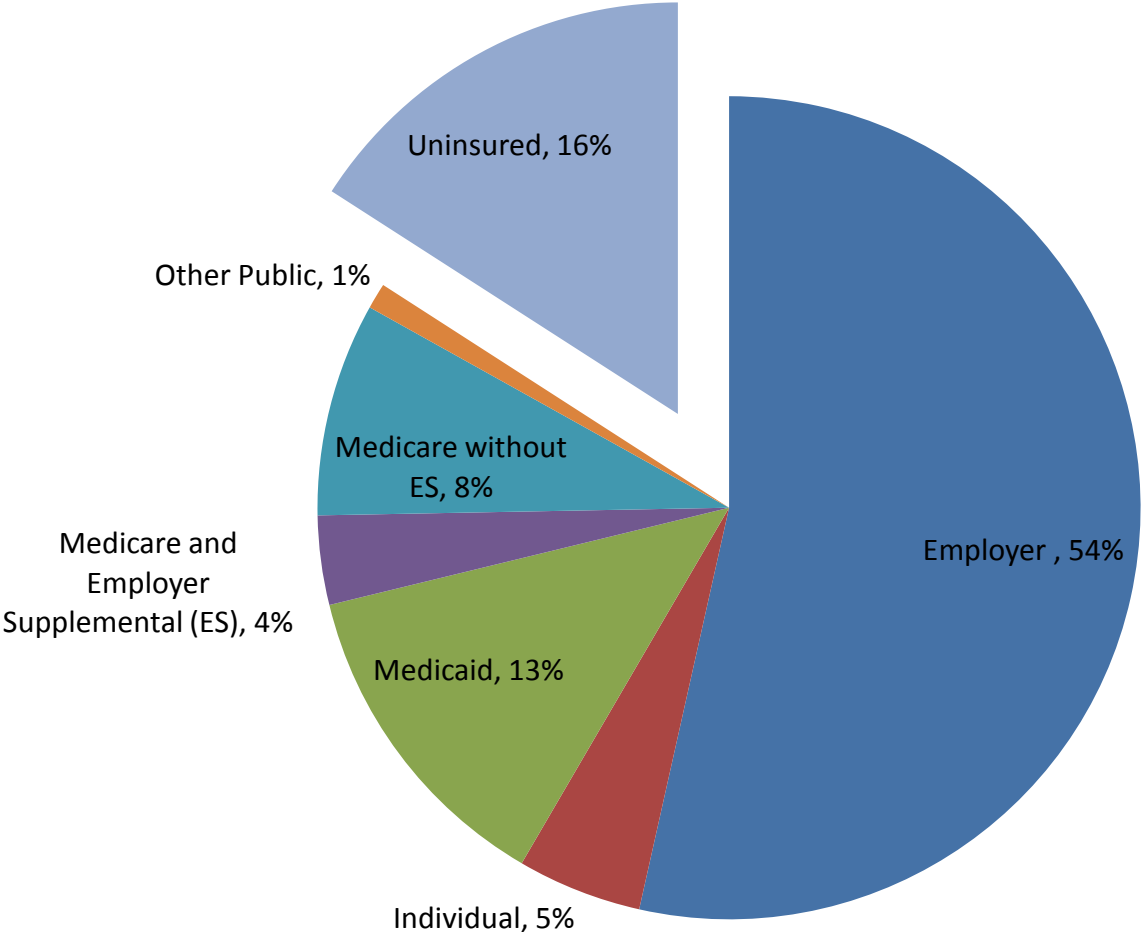
four of the alternatives produce very similar results in practice. They then suggest an efficient method for researchers to use when selecting estimators of health care costs. Researcher considering alternative models where the probability of use per se is not of interest would do well to start with the one-part GLM models.

Basu
(2009)

Single-Equation models: OLS regression for logarithmic or MLE estimation from Box-Cox transformations. Sophisticated Single Equation: One part generalized linear model (GLM) with a gamma distribution and a log link. Two-part models : Two-part generalized linear model (GLM) with a gamma distribution and a log link

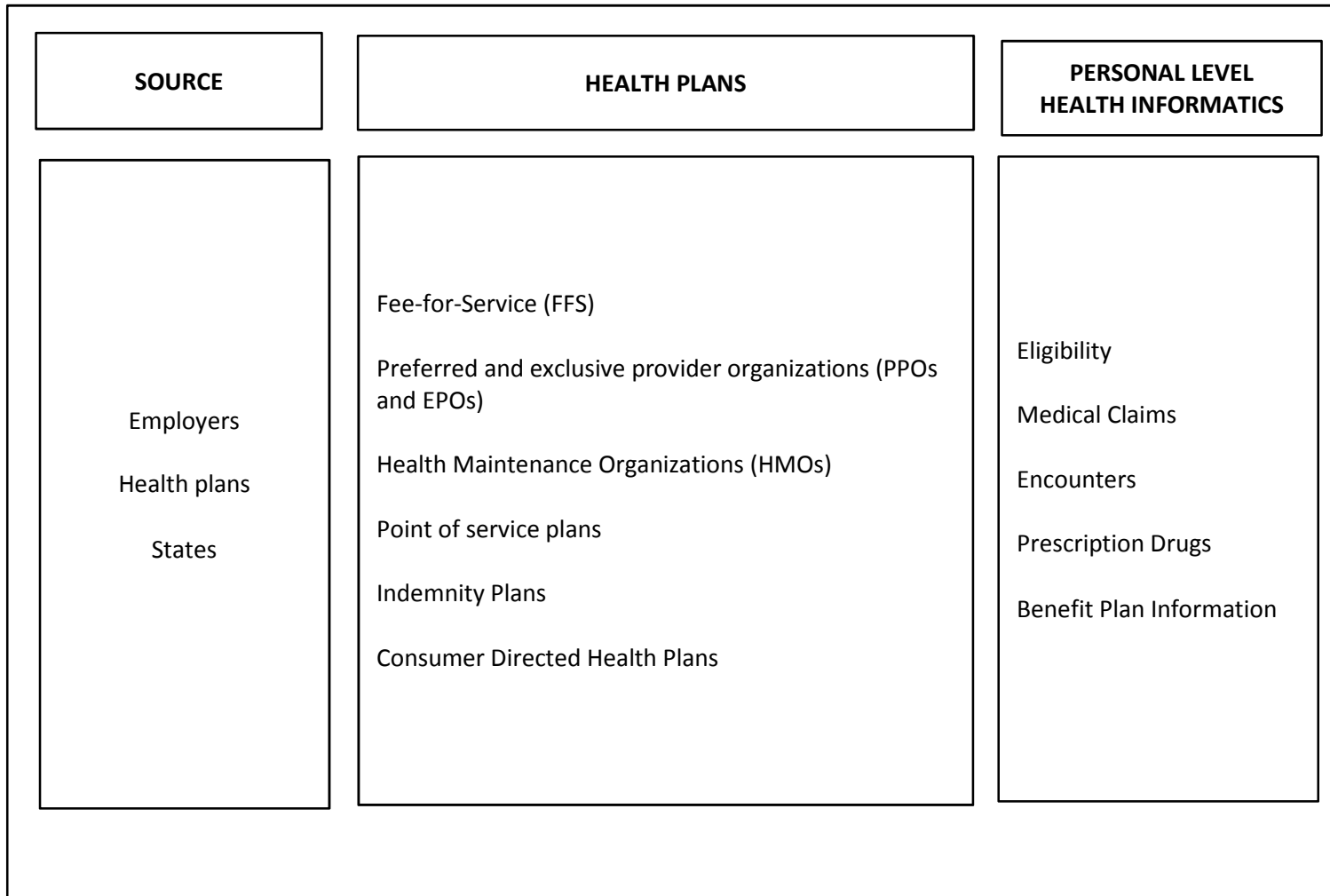
Given zero spenders and skewed positive expenditures data can be best handled by One part or two-part generalized linear model (GLM) with a gamma distribution and a log link. In two part model, use logit model to predict the probability of having any medical spending and then use a GLM model with a gamma distribution and a log link to estimate the level of expenditures, given positive spending.

Figure 1: Population distribution by Insurance Coverage - 2006



Notes: State Health Facts Online, The Henry J. Kaiser Foundation. US residents - 296 million

Figure 2: MarketScan Claims structure



Note : Source: Truven Health Analytics

Table 6: Relevant variables in MarketScan data

Demographic Variables	Enrollment data	Health Plan features	Inpatient claims and Outpatients claims	Drug Claims	Payment information
Enrollee identification	Date of enrollment	Plan type	Enrollee identification	Enrollee identification	Total payments
Age of patient			Date of admission	National Drug Code	
Patient birth year	Member days	Deductible amount	Date of discharge	Pharmacy id	Net Payments
Gender of patient			Length of stay	Date service incurred	
Relationship of patient to employee	Date of disenrollment	Copayment amount	Diagnosis related group	Therapeutic group	Payments to physicians
			Principal diagnosis code	Refill number	
Employment Status		Coordination of benefits amount	'Secondary diagnosis codes (up to 14)	Therapeutic class	Payments to hospitals
			Principal procedure code	Average Wholesale Price	
Employment classification			Secondary procedure codes (up to 14)	Coinsurance/ Copayment	Payments total admission
Industry			Place of service		
			Type of admission	Number of days supply	
Geographic location (state, zip code)			Provider ID	Deductible	
			Quality of services	Generic product identification	

Note : Source: Truven Health Analytics

Table 7. Summary Statistics for Sample Persons and Their Encounters/Claims, 2006

Characteristic	N	Percent
<i>Total Claims</i>	71,665,728	
Number of Claims by Type		
Inpatient	137,628	0.2
Outpatient	47,641,979	66.5
Drug	23,886,121	33.3
Mean (Median) Cost per Claim by Type		
Inpatient	\$14,134 (\$8,076)	
Outpatient	\$104 (\$37)	
Drug	\$83 (\$41)	
<i>Total Persons</i>	2,346,934	
Age		
<18	607,937	25.9
18-34	459,470	19.6
35-44	406,129	17.3
45-54	486,100	20.7
55-64	387,298	16.5
Female Gender	1,204,089	51.3
Region		
Northeast	280,951	12
North Central	619,047	26.4
South	1,070,411	45.6
West	357,558	15.2
Unknown	18,967	0.9
Mean (Median) Annual per Person Cost		
Total	\$3,788 (\$1,640)	
Inpatient	\$829 (\$474)	
Outpatient	\$2,118 (\$753)	
Drug	\$841 (\$414)	

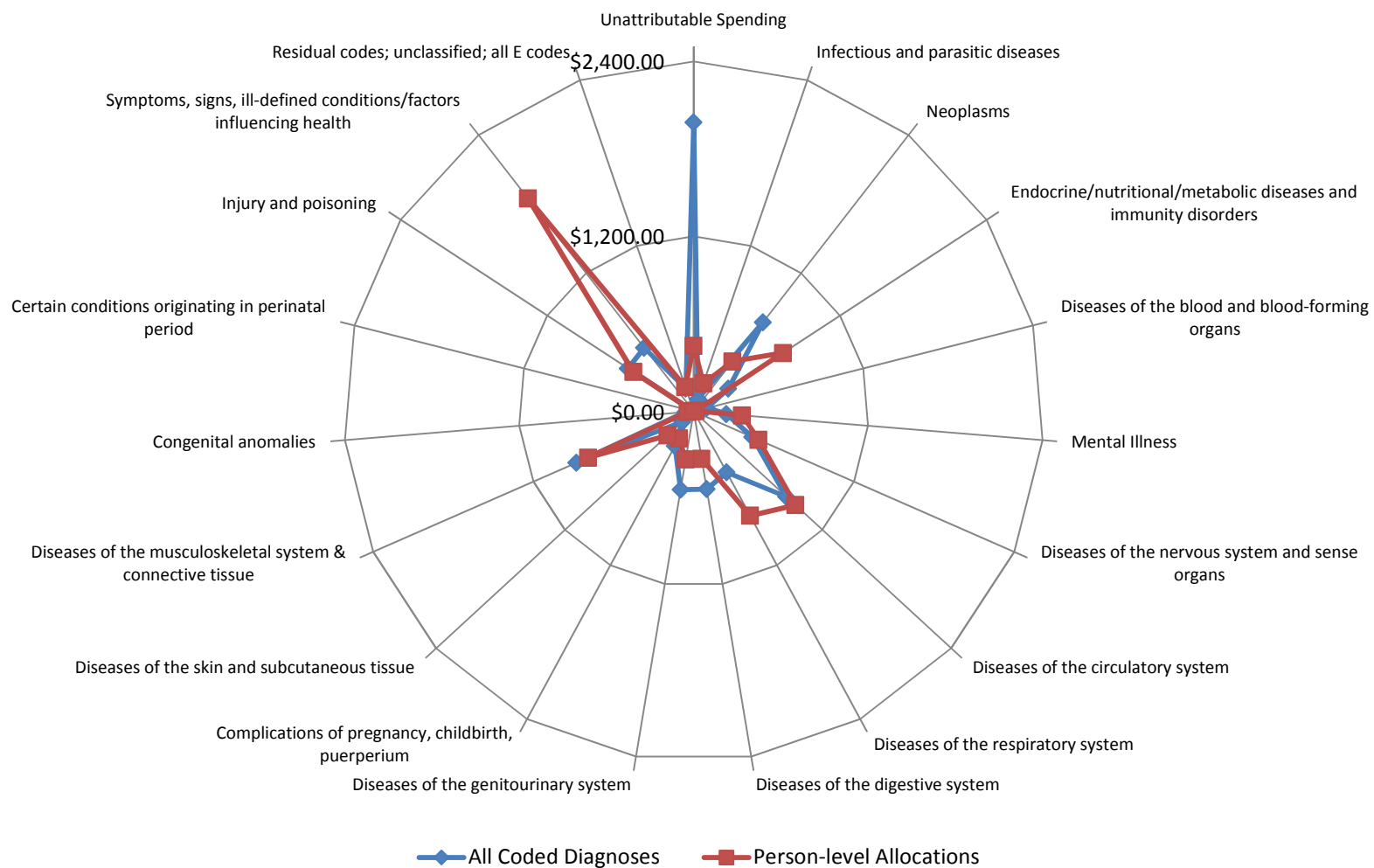
Notes: We restricted our analysis by randomly selecting approximately 3 million individuals in Market Scan data under the age of 65 with commercial insurance and prescription drug coverage in 2006. We excluded 0.58 million individuals with capitated insurance plans and dropped those with negative spending. The final analytic sample included 2.3 million individuals with 71.7 million claims totaling \$8.89 billion in annual spending (in 2006 U.S. dollars).

Table 8. Disease Spending Estimates by Method of Allocation

Condition Category	Encounter-level Allocations						Person-level Allocations		
	Principal Diagnosis Only			All Coded Diagnoses			Treated Prevalence	Cost per patient	Total Cost (millions)
	Treated Prevalence	Cost per patient	Total Cost (millions)	Treated Prevalence	Cost per patient	Total Cost (millions)			
<i>Unattributable Spending</i>			\$1,981.1			\$1,981.1			\$450.0
Infectious and parasitic diseases	18.1%	\$239	\$101.7	19.2%	\$200	\$90.4	19.2%	\$450	\$203.1
Neoplasms	10.4%	\$3,315	\$809.5	10.9%	\$3,019	\$774.6	10.9%	\$1,690	\$433.7
Endocrine/nutritional/metabolic diseases and immunity disorders	21.2%	\$533	\$264.8	23.0%	\$526	\$284.6	23.0%	\$1,353	\$731.8
Diseases of the blood and blood-forming organs	2.8%	\$1,077	\$71.8	3.4%	\$1,176	\$92.5	3.4%	\$195	\$15.4
Mental Illness	8.6%	\$997	\$201.9	10.1%	\$950	\$225.7	10.1%	\$1,403	\$333.3
Diseases of the nervous system and sense organs	24.6%	\$714	\$411.8	25.8%	\$728	\$439.9	25.8%	\$803	\$485.5
Diseases of the circulatory system	20.3%	\$1,806	\$859.3	21.6%	\$1,705	\$863.6	21.6%	\$1,873	\$949.0
Diseases of the respiratory system	35.2%	\$523	\$432.3	36.2%	\$560	\$475.8	36.2%	\$956	\$813.3
Diseases of the digestive system	14.0%	\$1,644	\$539.3	15.1%	\$1,524	\$540.6	15.1%	\$929	\$329.6
Diseases of the genitourinary system	20.7%	\$1,127	\$547.4	21.6%	\$1,075	\$544.5	21.6%	\$660	\$334.2
Complications of pregnancy, childbirth, puerperium	3.2%	\$3,865	\$292.4	3.4%	\$3,365	\$269.8	3.4%	\$2,645	\$212.1
Diseases of the skin and subcutaneous tissue	13.8%	\$348	\$112.6	14.4%	\$314	\$105.9	14.4%	\$721	\$243.3
Diseases of the musculoskeletal system & connective tissue	25.8%	\$1,554	\$941.8	26.8%	\$1,399	\$878.9	26.8%	\$1,258	\$790.5
Congenital anomalies	1.4%	\$1,666	\$56.2	1.6%	\$1,462	\$55.5	1.6%	\$1,090	\$41.4
Certain conditions originating in perinatal period	0.2%	\$1,271	\$6.9	0.4%	\$1,429	\$12.1	0.4%	\$513	\$4.3
Injury and poisoning	17.3%	\$1,454	\$590.0	17.8%	\$1,289	\$538.3	17.8%	\$1,182	\$493.8
Symptoms, signs, ill-defined conditions/factors influencing health	51.1%	\$459	\$551.5	52.5%	\$448	\$552.0	52.5%	\$1,502	\$1,849.3
Residual codes; unclassified; all E codes	7.8%	\$641	\$117.4	10.5%	\$668	\$163.8	10.5%	\$719	\$176.3
Total			\$8,889.79			\$8,889.79			\$8,889.79

Notes: The final analytic sample included 2.3 million individuals with 71.7 million claims totaling \$8.89 billion in annual spending (in 2006 U.S. dollars). We have attributed \$8.89 billion spending among 18 Broad ICD9 disease categories.

Figure 3: Total Cost (millions)



Notes: We have attributed \$8.89 billion spending among 18 Broad ICD9 disease categories. The biggest difference in attributable spending by the two methods is for “Symptoms, signs, ill-defined conditions/factors influencing health”

Table 9: Spending on Ten Most Expensive Diseases by Method

Disease	OVERALL COSTS (\$) by Method					
	1ry-Encounter		All-Encounter		Person	
	Rank	Dollars	Rank	Dollars	Rank	Dollars
Spondylosis, intervertebral disc and other back problems	1	389,715,873	1	357,920,863	5	304,271,176
Coronary atherosclerosis and other heart disease	2	196,579,627	2	182,455,430		124,485,308
Other connective tissue disease	3	150,535,660	3	166,567,806	7	200,708,175
Nonspecific chest pain	4	143,561,929	4	155,836,027		121,940,108
Osteoarthritis	5	133,035,281		96,306,494		18,174,807
Other and unspecified benign neoplasm	6	126,971,601	7	126,254,146		112,859,389
Cancer of breast	7	124,172,316	8	124,645,215		46,350,248
Abdominal pain	8	117,988,870	6	134,284,641	10	164,797,269
Residual codes; unclassified	9	117,230,458	5	153,625,332	9	170,533,590
Medical examination/evaluation	10	107,973,591		107,561,624	1	971,356,981
Other non-traumatic joint disorders		101,100,976	9	114,490,237		157,060,353
Other upper respiratory infections		99,226,131		96,231,728	3	399,125,448
Other screening for conditions (not mental or infectious)		92,026,348		85,270,130	6	289,369,276
Other lower respiratory disease		72,178,215	10	113,797,598		96,880,125
Essential hypertension		55,994,716		80,485,942	2	520,802,516
Disorders of lipid metabolism		41,044,961		47,001,777	4	379,037,069
Diabetes mellitus without complication		35,194,998		44,886,469	8	195,275,011
Overall Spending on:						
Top 10 conditions with method		1,607,765,205		1,629,877,294		3,595,276,513
All 17 Conditions in table (includes all top 10s)		2,104,531,550		2,187,621,458		4,273,026,852

Appendix 1 : CCS categories and ICD9 -CM codes for all 17 Conditions in Table 9

205 Spondylosis; intervertebral disc disorders; other back problems

7201 7202 72081 72089 7209 7210 7211 7212 7213 72141 72142 7215 7216 7217 7218 72190 72191 7220 72210 72211 7222 72230
72231 72232 72239 7224 72251 72252 7226 72270 72271 72272 72273 72280 72281 72282 72283 72290 72291 72292 72293 7230
7231 7232 7233 7234 7235 7236 7237 7238 7239 72400 72401 72402 72403 72409 7241 7242 7243 7244 7245 7246 72470 72471
72479 7248 7249

101 Coronary atherosclerosis and other heart disease

4110 4111 4118 41181 41189 412 4130 4131 4139 4140 41400 41401 41406 4142 4143 4144 4148 4149 V4581 V4582

211 Other connective tissue disease

32752 56731 7105 725 7260 72610 72611 72612 72613 72619 7262 72630 72631 72632 72633 72639 7264 7265 72660 72661 72662
72663 72664 72665 72669 72670 72671 72672 72673 72679 7268 72690 72691 72700 72701 72702 72703 72704 72705 72706 72709
7272 7273 72740 72741 72742 72743 72749 72750 72751 72759 72760 72761 72762 72763 72764 72765 72766 72767 72768 72769
72781 72782 72783 72789 7279 7280 72810 72811 72812 72813 72819 7282 7283 7284 7285 7286 72871 72879 72881 72882 72883
72884 72885 72886 72887 72888 72889 7289 7290 7291 7292 72930 72931 72939 7294 7295 7296 72971 72972 72973 72979 72981
72982 72989 7299 72990 72991 72992 72999 7819 78191 78192 78194 78199 7937 V135 V1359 V436 V4360 V4361 V4362 V4363
V4364 V4365 V4366 V4369 V437 V454 V481 V482 V483 V490 V491 V492 V495 V4960 V4961 V4962 V4963 V4964 V4965 V4966 V4967
V4970 V4971 V4972 V4973 V4974 V4975 V4976 V4977 V537

102 Nonspecific chest pain

78650 78651 78659

203 Osteoarthritis

71500 71504 71509 71510 71511 71512 71513 71514 71515 71516 71517 71518 71520 71521 71522 71523 71524 71525 71526 71527
71528 71530 71531 71532 71533 71534 71535 71536 71537 71538 71580 71589 71590 71591 71592 71593 71594 71595 71596 71597
71598 V134

47 Other and unspecified benign neoplasm

20940 20941 20942 20943 20950 20951 20952 20953 20954 20955 20956 20957 20960 20961 20962 20963 20964 20965 20966 20967
20969 2100 2101 2102 2103 2104 2105 2106 2107 2108 2109 2110 2111 2112 2113 2114 2115 2116 2117 2118 2119 2120 2121 2122
2123 2124 2125 2126 2127 2128 2129 2130 2131 2132 2133 2134 2135 2136 2137 2138 2139 2140 2141 2142 2143 2144 2148 2149
2150 2152 2153 2154 2155 2156 2157 2158 2159 2160 2161 2162 2163 2164 2165 2166 2167 2168 2169 217 220 2210 2211 2212 2218
2219 2220 2221 2222 2223 2224 2228 2229 2230 2231 2232 2233 22381 22389 2239 2240 2241 2242 2243 2244 2245 2246 2247 2248
2249 2250 2251 2252 2253 2254 2258 2259 226 2270 2271 2273 2274 2275 2276 2278 2279 22800 22801 22802 22803 22804 22809
2281 2290 2298 2299 V1272

24 Cancer of breast

1740 1741 1742 1743 1744 1745 1746 1748 1749 1750 1759 2330 V103

251 Abdominal pain

7890 78900 78901 78902 78903 78904 78905 78906 78907 78909 78960 78961 78962 78963 78964 78965 78966 78967 78969

259 Residual codes; unclassified

3020 32700 32701 32709 32710 32711 32712 32713 32714 32719 32720 32721 32722 32723 32724 32725 32726 32727 32729 32740
32741 32742 32743 32744 32749 32751 32759 3278 78002 7801 78050 78051 78052 78053 78054 78055 78056 78057 78058 78059
78064 78065 7809 78093 78094 78095 78096 78097 78099 7815 7816 7823 78261 78262 7828 7829 7830 7836 7842 7901 7906 7909
79091 79092 79093 79094 79095 79099 7932 7939 79399 7949 79581 79582 79589 7963 7964 7965 7966 7969 7980 7981 7982 7989
7992 79921 79922 79923 79924 79925 79929 7993 7998 79981 79982 79989 7999 V070 V072 V073 V0731 V0739 V0751 V0752 V0759
V078 V079 V131 V138 V1389 V139 V152 V1521 V1522 V1529 V153 V1581 V1584 V1585 V1586 V1587 V1589 V159 V160 V161 V162
V163 V164 V1640 V1641 V1642 V1643 V1649 V165 V1651 V1652 V1659 V166 V167 V168 V169 V170 V171 V172 V173 V174 V1741
V1749 V175 V176 V177 V178 V1781 V1789 V180 V181 V1811 V1819 V182 V183 V184 V185 V1851 V1859 V186 V1861 V1869 V187
V188 V189 V190 V191 V1911 V1919 V192 V193 V194 V195 V196 V197 V198 V210 V211 V21 218 V219 V418 V419 V428 V4281
V4282 V4283 V4284 V4289 V429 V438 V4381 V4382 V4383 V4389 V447 V448 V449 V4571 V4572 V4573 V4574 V4575 V4576 V4577
V4578 V4579 V4583 V4584 V4586 V4587 V4588 V4589 V460 V463 V468 V469 V470 V471 V472 V479 V480 V488 V489 V498 V4981
V4982 V4983 V4984 V4986 V4987 V4989 V499 V500 V501 V503 V5041 V5042 V5049 V508 V509 V590 V5901 V5902 V5909 V591 V592
V593 V594 V595 V596 V5970 V5971 V5972 V5973 V5974 V598 V599 V640 V6400 V6401 V6402 V6403 V6404 V6405 V6406 V6407
V6408 V6409 V641 V642 V643 V644 V6441 V6442 V6443 V690 V691 V692 V693 V694 V695 V698 V699 V8301 V8302 V8381 V8389
V8401 V8402 V8403 V8404 V8409 V848 V8481 V8489 V851 V8552 V860 V861 V8701 V8702 V8709 V8711 V8712 V8719 V872 V8731
V8732 V8739 V8741 V8742 V8743 V8744 V8745 V8746 V8749 V8801 V8802 V8803 V8811 V8812 V8901 V8902 V8903 V8904 V8905
V8909

256 Medical examination/evaluation

V290 V291 V292 V293 V298 V299 V6801 V6809 V700 V703 V704 V705 V706 V707 V708 V709 V718 V719 V7231 V7232 V725 V726
V7260 V7261 V7262 V7263 V7269 V728 V7281 V7282 V7283 V7284 V7285 V7286 V729

204 Other non-traumatic joint disorders

7130 7131 7132 7133 7134 7135 7136 7137 7138 71600 71601 71602 71603 71604 71605 71606 71607 71608 71609 71620 71621
71622 71623 71624 71625 71626 71627 71628 71629 71630 71631 71632 71633 71634 71635 71636 71637 71638 71639 71640 71641
71642 71643 71644 71645 71646 71647 71648 71649 71650 71651 71652 71653 71654 71655 71656 71657 71658 71659 71660 71661
71662 71663 71664 71665 71666 71667 71668 71680 71681 71682 71683 71684 71685 71686 71687 71688 71689 71690 71691 71692
71693 71694 71695 71696 71697 71698 71699 71810 71811 71812 71813 71814 71815 71817 71818 71819 71820 71821 71822 71823
71824 71825 71826 71827 71828 71829 71850 71851 71852 71853 71854 71855 71856 71857 71858 71859 71860 71865 71870 71871
71872 71873 71874 71875 71876 71877 71878 71879 71880 71881 71882 71883 71884 71885 71886 71887 71888 71889 71890 71891
71892 71893 71894 71895 71897 71898 71899 71900 71901 71902 71903 71904 71905 71906 71907 71908 71909 71910 71911 71912
71913 71914 71915 71916 71917 71918 71919 71920 71921 71922 72923 71924 71925 71926 71927 71928 71929 71930 71931 71932
71933 71934 71935 71936 71937 71938 71939 71940 71941 71942 71943 71944 71945 71946 71947 71948 71949 71950 71951 71952
71953 71954 71955 71956 71957 71958 71959 71960 71961 71962 71963 71964 71965 71966 71967 71968 71969 7197 71970 71975
71976 71977 71978 71979 71980 71981 71982 71983 71984 71985 71986 71987 71988 71989 71990 71991 71992 71993 71994 71995
71996 71997 71998 71999

126 Other upper respiratory infections

0320 0321 0322 0323 0340 460 4610 4611 4612 4613 4618 4619 462 4640 46400 46401 46410 46411 46420 46421 46430 46431 4644
46450 46451 4650 4658 4659 4730 4731 4732 4733 4738 4739 78491

10 Immunizations and screening for infectious disease

7955 79551 79552 7956 V010 V011 V012 V013 V014 V015 V016 V017 V0171 V0179 V018 V0181 V0182 V0183 V0184 V0189 V019
V020 V021 V022 V023 V024 V025 V0251 V0252 V0253 V0254 V0259 V026 V0260 V0261 V0262 V0269 V027 V028 V029 V030 V031
V032 V033 V034 V035 V036 V037 V038 V0381 V0382 V0389 V039 V040 V041 V042 V043 V044 V045 V046 V047 V048 V0481 V0482
V0489 V050 V051 V052 V053 V054 V058 V059 V060 V061 V062 V063 V064 V065 V066 V068 V069 V286 V712 V7182 V7183 V730
V731 V732 V733 V734 V735 V736 V738 V7381 V7388 V7389 V739 V7398 V7399 V740 V741 V742 V743 V744 V745 V746 V748 V749
V750 V751 V752 V753 V754 V755 V756 V757 V758 V759 79579

133 Other lower respiratory disease

5131 514 515 5160 5161 5162 5163 51630 51631 51632 51633 51634 51635 51636 51637 5164 5165 51661 51662 51663 51664 51669
5168 5169 5172 5178 5183 5184 51889 5194 5198 5199 7825 78600 78601 78602 78603 78604 78605 78606 78607 78609 7862 7863
78630 78631 78639 7864 78652 7866 7867 7868 7869 7931 79311 79319 7942 V126 V1260 V1261 V1269 V426

98 Essential hypertension

4011 4019

53 Disorders of lipid metabolism

2720 2721 2722 2723 2724

49 Diabetes mellitus without complication

24900 25000 25001 7902 79021 79022 79029 7915 7916 V4585 V5391 V6546
