NBER WORKING PAPER SERIES

CALCULATING DISEASE-BASED MEDICAL CARE EXPENDITURE INDEXES FOR MEDICARE BENEFICIARIES: A COMPARISON OF METHOD AND DATA CHOICES

Anne E. Hall
Tina Highfill

Working Paper 19720
http://www.nber.org/papers/w19720

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2013

We would like to thank Ana Aizcorbe and Abe Dunn for helpful comments and advice. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Calculating Disease-Based Medical Care Expenditure Indexes for Medicare Beneficiaries:
A Comparison of Method and Data Choices
Anne E. Hall and Tina Highfill
NBER Working Paper No. 19720
December 2013
JEL No. I1

ABSTRACT

Disease-based medical care expenditure indexes are currently of interest to measurement economists. In this paper, using two data sources and two different methods for calculating disease-based expenditure indexes for the Medicare population, we establish some results that will help guide policymakers in choosing indexes for this population. First, we find that the two methods we examine (primary diagnosis and a regression-based method) produce the same results for the aggregate index and have a moderate level of agreement in which diseases contribute the most to growth in per capita health-care spending. Since the primary diagnosis method is preferable because of its transparency, this result implies that we may use the regression-based method when the data is not suitable for the primary diagnosis method without a great loss of accuracy. Second, we find that the two data sources, the Medicare Current Beneficiary Survey and the Medical Expenditure Panel Survey, produce very similar results in the aggregate but there is some evidence that they treat chronic illnesses differently. As the MCBS has a larger sample and more comprehensive coverage of Medicare beneficiaries than the MEPS, it seems that a regression-based expenditure index based on the MCBS is overall preferable for fee-for-service Medicare beneficiaries.

Anne E. Hall
Bureau of Economic Analysis
anne.hall@bea.gov

Tina Highfill
Bureau of Economic Analysis
tina.highfill@bea.gov

Introduction

Recent research has shown the effect of new methods on the measurement of inflation in health care. Specifically, papers such as Aizcorbe and Nestoriak (2011), Dunn et al. (2010), and our previous work (Hall and Highfill 2013) followed the recommendation of the Committee on National Statistics of the National Research Council that statistical agencies investigate methods and datasets for measuring health-care expenditure by disease and for creating new price indexes based on the costs of episodes of treatment of specific illnesses. These papers found average annual growth rates for health-care inflation that ranged from 2.9 percent to 6.9 percent. However, the papers used different datasets on different populations and employed different methods for splitting up health-care expenditure by disease, making it difficult to analyze why their results varied so much.

In this paper, we are therefore following up on our previous work by comparing medical expenditure indexes for the Medicare population calculated from different datasets and using different methods. The Medicare program plays a very large role in the US healthcare system: it covered 50.8 million people in calendar year 2012 (42.2 million aged and 8.6 million disabled), and its payments comprised 22.9% of the total federal budget and 12.6% of total national health-care spending in 2012. In addition, the Congressional Budget Office expects the aging of the population and rising health-care costs, through their effect on Medicare spending, to be major factors in the projected increase in the deficit later this decade (CBO 2013).

We compare two major data sources on Medicare beneficiaries: the Medicare Current Beneficiary Survey (MCBS) and the Medical Expenditure Panel Survey (MEPS). Both are surveys conducted by the Department of Health and Human Services; the MCBS surveys Medicare beneficiaries exclusively while the MEPS surveys US residents living in the community about their health care and spending. We will also compare two methods for calculating medical expenditure indexes: the primary diagnosis method, which simply assigns spending to the illness associated with the diagnosis code (or first diagnosis code in the case of multiple diagnosis codes) of each claim or survey-collected medical event; and a regression-based approach which regresses individual annual health-care spending on dummy variables for

medical conditions each beneficiary is diagnosed with and divides up each beneficiary's health-care spending on that basis.

We find, on the whole, that the expenditure indexes give about the same average annual growth rate of about 3 percent whichever dataset or method is used. In terms of method, while the primary diagnosis method is a priori preferable as the most transparent method, it has the stringent requirement that every claim or event have a diagnosis attached to it, a requirement which is only met by the MEPS. The MCBS, however, is generally a preferable dataset for studying Medicare beneficiaries since the MEPS has a smaller sample size than the MCBS, does not include nursing home residents, and appears to have a fair amount of underreporting of spending when it is compared directly with the MCBS. In our comparison of the two methods when we apply them to the same samples from the MEPS and from the MCBS (without drug spending), we find that both methods produce about the same average annual growth rate of the overall expenditure index. This result suggests that the regression-based method may be applied to the MCBS without too great a loss of accuracy.

We also compare indexes calculated from the MCBS and the MEPS using the same method. When drug spending is included and we use a regression-based method, the two datasets produce about the same results in the aggregate. When drug spending is dropped, expenditure indexes created from the MEPS grow about a percentage point faster than the indexes based on the MCBS. We show some evidence that this difference may relate to the way MEPS treats chronic illnesses.

By necessity, our comparison focuses on the potential indexes for Medicare beneficiaries enrolled in fee-for-service (FFS) Medicare, for whom the most complete data exists. Medicare private-plan enrollees are, however, making up an increasing share of Medicare enrollment but the data for them is much less comprehensive. In the MCBS, the only source for their spending, diagnoses and events is the in-person survey but as we shall see, the survey asks about a limited set of diagnoses and some important ones would be omitted if we based their expenditure index on the MCBS survey. However, the Medicare private plan sample in the MEPS is quite small, only about 300 to 400 beneficiaries per year.

Methods

The problem of creating expenditure indexes by disease for health care comes down to how to split up individuals' health-care expenditure across the diseases they are diagnosed with. Several different methods have been proposed and used in prior papers. It should be noted that economic theory is no guide on methodology in this area and there is no true way of validating any of the methods. Brief descriptions of the methods available to us follow here and the results of the previous papers are summarized in Table 1.

Regression-based: This method is discussed in detail in Hall and Highfill (2013). Briefly, the log of individual total medical spending on all conditions is regressed on dummy variables for the medical conditions in the dataset. Following Trogdon (2008), the resulting coefficients are used to calculate shares of individual medical spending assigned to each condition the individual is diagnosed with. This method has the advantage that it does not make huge demands of the data, compared to some of the other methods discussed here: it only requires individual annual medical spending and dummy variables for whether or not the individual was diagnosed with a certain condition that year. The primary disadvantage is that the method of assignment of spending to particular diagnoses is not very transparent and not based on any theory or model that relates health-care spending to conditions diagnosed.

Primary diagnosis: This method is used in Aizcorbe et al. (2011). With this method, the spending attached to a claim or medical event is assigned to the diagnosis or the first diagnosis, if there are multiple diagnoses, attached to that claim or event. In some sense, the primary diagnosis method is preferable to all others because the connection between spending and diagnosis is transparent. This method requires, however, that every claim or event have a diagnosis attached to it. Other than the MEPS, the datasets being considered for the Medicare population do not have a diagnosis attached to every claim or event; the Medicare Part D claims do not have diagnoses and the survey events collected by the Medicare Current Beneficiary Survey that are not in the Medicare claims (such as drug events and all medical events for Medicare private plan enrollees) do not have diagnoses attached to them either.

Commercial grouper: Several private companies have developed commercial software for grouping medical spending by episode of illness based on clinical knowledge. These packages are used in Aizcorbe and Nestoriak (2011), Dunn et al. (2010), and Aizcorbe et al. (2011) to create expenditure indexes for medical care. Aizcorbe and Nestoriak (2011) use one on medical claims data from private employer-sponsored health insurance plans collected by Pharmetrics, Inc., and Dunn et al. (2010) use one on the MarketScan database, a similar dataset of medical claims collected by Thomson Reuters. Aizcorbe et al. (2011) use a commercial grouper on the MEPS data. The main advantage of these packages over the primary diagnosis method is that they are able to assign spending associated with claims, such as drug claims, that do not have a diagnosis attached directly to them. In addition, unlike the annual regression-based method, they are able to separate out multiple episodes of the same illness occurring in one year. Depending on the package, they are also able to assign a severity level to the illness. However, their methods are proprietary and therefore completely opaque to the economist using them. We will not be considering groupers in our comparisons in this paper but may study them in future work.

Table 1: Previous papers calculating medical care expenditure indexes					
Paper	**Dataset**	**Population**	**Years**	**Method**	**Results (AAGR of price index in percentage points)**
Aizcorbe and Nestoriak 2011	Pharmetrics, Inc. (medical claims from private employer-sponsored health insurance plans)	Beneficiaries of private employer-sponsored insurance	2003-2005	Commercial grouper (Symmetry)	3.7
Dunn et al. 2010	MarketScan (medical claims from private employer-sponsored health insurance plans)	Beneficiaries of private employer-sponsored insurance	2003-2007	Commercial grouper (Symmetry)	3.6
Aizcorbe et al. 2011	Medical Expenditure Panel Survey (MEPS)	People living in the community	2001-2005	Primary diagnosis, proportional diagnosis, commercial grouper (Thomson Reuters)	6.9, 6.6, 6.8
Hall and Highfill 2013	Medicare Current Beneficiary Survey	Medicare beneficiaries	2001-2005	Regression-based	5.8

Datasets

Table 2 summarizes the features of datasets available for calculating medical care expenditure indexes for Medicare beneficiaries. In general, to create a medical care expenditure index, we need variables for total spending and for diagnoses at at least an annual level.

Table 2: Summary of datasets covering Medicare beneficiaries

Dataset	Coverage	Annual sample size of Medicare beneficiaries	Data available for FFS beneficiaries	Data available for private plan enrollees
Medicare Current Beneficiary Survey (MCBS)	All Medicare beneficiaries	12,000	Annual demographic and conditions survey, all medical events and spending, Medicare Part A (hospital) and Part B (physician) claims	Annual demographic and conditions survey, all medical events and spending
Medicare claims	FFS Medicare beneficiaries	~2 million	Part A and Part B claims; Part D (pharmaceutical) claims for about 50-60% of 5% sample	n/a
Medical Expenditure Panel Survey (MEPS)	Medicare beneficiaries living in the community	4,600	All medical events and spending, with diagnoses attached (collected by survey)	All medical events and spending, with diagnoses attached (collected by survey)

Medicare Current Beneficiary Survey (MCBS): The MCBS is a survey of the demographics, diagnosed conditions, health status, and total medical spending of a representative sample of Medicare beneficiaries. It is conducted by the Center for Medicare and Medicaid Services, the agency that operates Medicare. As it samples from the universe of Medicare beneficiaries, it includes both FFS Medicare beneficiaries and those enrolled in Medicare private plans, and

both beneficiaries residing in the community and in nursing homes. The medical conditions portion of the survey takes place once a year, towards the end of the year, and in it, the respondent is asked whether they have been told by a doctor if they have each of about 30 conditions. Their spending and medical events are also collected directly from the respondent on a regular basis. For FFS beneficiaries, the Part A and Part B claims with dollar amounts and diagnosis codes are also attached to the survey so there are two sources of diagnoses and spending for these beneficiaries. MCBS reconciles the orally reported events and the claims so that spending and events are not duplicated in the final version of the survey. For private plan enrollees, the only source of information is the spending, events, and diagnoses reported in the in-person survey.

Medicare claims: Part A (hospital) and Part B (doctor) claims are available for a 5% random sample of Medicare beneficiaries from CMS for research purposes. In addition, starting in 2006, a sample of Part D claims are available for those 50-60% of FFS beneficiaries in the 5% who are on Part D. In this paper, rather than showing calculations from the full 5% sample, we will evaluate Medicare claims data by using the Medicare claims data tied to the MCBS.

Medical Expenditure Panel Survey (MEPS): The MEPS is a nationally representative survey of healthcare coverage, utilization, and expenditures for the civilian non-institutionalized U.S. population. It is conducted by the Department of Health and Human Services' Agency for Healthcare Research and Quality (AHRQ). The survey sample is drawn from the respondents of the prior year's National Health Interview Survey (NHIS) and includes both FFS and private plan Medicare beneficiaries living in the community. Using an overlapping panel design, each household is surveyed over the course of two years in five rounds of interviews. The family member most knowledgeable about the entire household's health and health care use is interviewed. Observations are collected and reported for every medical event and may contain up to four diagnoses each. The MEPS also collects data from a sample of respondents' providers to verify use of services, charges and sources of payments, and diagnoses.

One disadvantage of the MEPS is that there is apparently quite a bit of underreporting of spending and events. As Table 3 shows, when we compare average spending and the

Table 3: Comparison of spending in MCBS and MEPS

	Average spending			Median spending		Standard deviation		Coefficient of variation		Skewness coefficient	
	MCBS FFS	MCBS FFS community only	MEPS FFS	MCBS FFS community only	MEPS FFS	MCBS community only	MEPS FFS	MCBS FFS community only	MEPS FFS	MCBS FFS community only	MEPS FFS
2001	$8,698	$7,701	$5,846	$3,723	$2,570	$ 760,036	$ 1,054,085	98.7	180.3	10.1	7.3
2002	$9,237	$8,309	$6,676	$4,169	$3,104	$ 759,332	$ 1,049,874	91.4	157.3	6.5	5.5
2003	$10,027	$8,902	$7,138	$4,483	$3,118	$ 881,963	$ 1,134,714	99.1	159.0	13.9	4.8
2004	$10,759	$9,741	$7,688	$4,846	$3,577	$ 1,054,373	$ 1,224,369	108.2	159.3	18.1	5.7
2005	$11,737	$10,638	$8,001	$5,701	$3,599	$ 1,060,720	$ 1,405,882	99.7	175.7	10.3	7.8
2006	$12,168	$10,938	$7,899	$5,954	$3,568	$ 1,065,798	$ 1,483,608	97.4	187.8	14.3	11.3
2007	$12,671	$11,433	$8,566	$6,220	$3,699	$ 1,111,613	$ 1,479,498	97.2	172.7	12.0	5.2
2008	$13,371	$12,034	$8,335	$6,216	$4,062	$ 1,373,493	$ 1,334,976	114.1	160.2	14.3	4.3
2009	$13,198	$11,855	$9,089	$6,450	$3,971	$ 1,108,157	$ 1,540,932	93.5	169.5	5.2	5.6

distribution of spending of FFS Medicare beneficiaries residing in the community in the MCBS and the MEPS, average spending in the MEPS is only about 70-80% of what is reported in MCBS. The median of spending in the MEPS is about 60-70% of that of the MCBS and while the coefficient of variation is higher in the MEPS (at about 160-180 compared with a coefficient of variation of about 100 in the MCBS), the skewness coefficient is higher in the MCBS, implying that the MCBS captures more of the higher end of the distribution of spending and therefore that the MEPS may be missing the highest spenders.

Prevalence

As discussed in the introduction, the effective goal of creating a medical care expenditure index is to divide up health-care spending into prices and quantities, where the prices are the average amount of money spent to treat illnesses and quantities are the number of people treated for those illnesses. Total health-care spending in a population can therefore be expressed as a sum over illnesses:

$$\$ = \sum_{j=1}^{J} \bar{P}_j N_j$$

where j indexes illnesses, P_j is the average amount spent to treat illness j and N_j is the number of people treated for illness j.

P_j is the object of concern when we are calculating a price index; however, the data we use have total spending and N_j as given and we must infer P_j with one of the various methods that will be discussed further down. Differences in treated prevalence between datasets will therefore lead to different results in the price index. It is therefore important to note that the survey portion of the MCBS, the claims portion of the MCBS, and the MEPS all have different prevalences for comparable illnesses. The differences may partly result from the different ways in which the diagnoses are collected: the community survey portion of the MCBS and the MEPS rely on patient-reported diagnoses, while the institutional survey portion of the MCBS (which is only collected for nursing home residents) and the Medicare claims diagnoses are reported by health-care practitioners.

For the FFS beneficiaries in the MCBS, we have both sets of diagnoses for the same beneficiaries and can examine the degree of agreement between them. Table 4 reports prevalences in 2001 from the MCBS survey and from the MCBS claims, as well as the percent of beneficiaries who are indicated as having an illness in both the survey and the claims. (For how the survey diagnoses were translated into ICD-9 codes for purposes of comparison with the claims, see Appendix Table 1 in Hall and Highfill 2013.) First, note that, in general, claims-based prevalence is higher than survey-based prevalence. It is difficult to know, however, whether this is because providers are upcoding beneficiaries' illnesses or because beneficiaries are not remembering all of their own diagnoses in the annual survey. There is also quite a bit of disagreement between the two sources of diagnoses and the agreement rate is often less than half of either the survey-based prevalence or the claims-based prevalence. When weighted by prevalence, the overlap rate divided by the claims-based prevalence averages about 40 percent and the overlap rate divided by the survey-based prevalence averages about 62 percent.

One possible cause of disagreement in prevalence between the survey and the claims may come from the lack of drug-related diagnoses in the claims data. In other words, beneficiaries may have chronic illnesses for which they are taking drugs regularly (and which they report on the survey) but these illnesses are failing to get recorded by their health-care providers. However, some of the highest rates of agreement between survey and claims are in illnesses for which this issue would most be of concern. There are five chronic illnesses for whom the share of spending on pharmaceuticals is over 50 percent of total spending on the illness: diabetes, mental/psychiatric disorder, Alzheimers/dementia, osteoporosis, and hypertension. Agreement rates between survey-based and claims-based prevalence are in fact noticeably higher for this group: the overlap rate divided by the claims-based prevalence has a weighted average of 63 percent and the overlap rate divided by the survey-based prevalence has a weighted average of 72 percent. It seems possible therefore that the act of taking a daily drug for an illness may actually improve survey respondents' knowledge and memory of what diagnoses they have.

Tables 5 and 6 illustrate a related problem with prevalences in the MEPS during this period. Table 5 compares the survey-based and claims-based prevalences of the MCBS in 2001 and

Table 4: Prevalence in survey portion and claims portion of MCBS in 2001

Condition	Survey Prevalence	Claim Prevalence	Overlap Prevalence
Hardening of arteries/arteriosclerotic heart disease	10.9%	20.4%	5.1%
Myocardial infarction/Heart attack	2.6%	2.1%	0.8%
Angina/CHD	3.8%	15.6%	2.4%
Other heart conditions and valve problem	5.6%	22.1%	3.2%
Congestive heart failure	3.3%	12.5%	2.4%
Heart rhythm problem	7.5%	19.3%	4.7%
Stroke/transient ischemic attack (TIA)	3.5%	8.1%	2.1%
Skin cancer	4.9%	5.0%	2.5%
Lung cancer	0.2%	0.9%	0.2%
Colon cancer	0.4%	1.5%	0.3%
Breast cancer	0.6%	2.4%	0.4%
Prostate cancer	0.6%	3.5%	0.5%
Other cancer	1.2%	13.6%	0.8%
Diabetes	19.3%	22.6%	16.9%
Arthritis	23.1%	24.8%	9.1%
Mental/psychiatric disorder (excl. Alzheimers/dementia)	12.2%	16.2%	7.0%
Alzheimers/dementia	5.0%	6.9%	3.3%
Osteoporosis	17.5%	11.1%	6.8%
Hypertension	44.4%	56.2%	37.2%
Broken hip	1.0%	1.4%	0.5%
Parkinsons	1.6%	1.4%	1.0%
Emphysema/asthma/chronic obstructive pulmonary disease (COPD)	15.5%	19.5%	9.5%
Paralysis in past year	1.8%	1.7%	0.5%
Mental retardation (excl. Alzheimers/dementia)	3.1%	0.9%	0.7%
Renal failure	0.8%	4.0%	0.8%

Notes: 1. Refer to Appendix Table 1 in Hall and Highfill 2013 to see how illnesses from the MCBS survey were translated into ICD-9 codes to compare with the claims data.

2. Prevalences are weighted by the MCBS survey weights.

Table 5: Comparison of prevalences across MCBS and MEPS

	2001			2009			AAGR in prevalence		
Condition	**MCBS Survey**	**MCBS Claims**	**MEPS**	**MCBS Survey**	**MCBS Claims**	**MEPS**	**MCBS Survey**	**MCBS Claims**	**MEPS**
Hardening of arteries/arteriosclerotic heart disease + Angina/CHD	13.4%	25.0%	4.3%	12.0%	24.1%	18.0%	-1.4%	-0.4%	29.5%
Lung cancer	0.2%	0.9%	0.4%	0.4%	1.1%	0.4%	12.2%	3.5%	16.5%
Colon cancer	0.4%	1.5%	0.5%	0.4%	0.9%	0.8%	2.8%	-5.5%	13.6%
Breast cancer	0.6%	2.4%	1.6%	0.7%	2.7%	2.5%	2.9%	1.5%	8.1%
Prostate cancer	0.6%	3.5%	1.8%	0.7%	4.0%	3.0%	2.4%	1.8%	11.4%
Other cancer	*1.2%*	*13.6%*	4.9%	1.4%	15.0%	5.8%	3.0%	1.3%	2.5%
Diabetes	19.3%	22.6%	16.0%	23.4%	31.0%	21.8%	2.5%	4.0%	4.1%
Arthritis	23.1%	24.8%	3.4%	19.7%	25.9%	15.3%	-1.3%	0.6%	30.2%
Mental/psychiatric disorder (excl. Alzheimers/dementia)	12.2%	16.2%	13.7%	17.9%	21.0%	18.9%	6.0%	3.4%	4.4%
Alzheimers/dementia	5.0%	6.9%	1.8%	6.7%	7.9%	2.5%	4.0%	1.8%	7.9%
Osteoporosis	17.5%	11.1%	5.5%	21.9%	12.0%	5.9%	2.9%	1.1%	2.1%
Hypertension	44.4%	56.2%	43.0%	49.8%	65.7%	63.2%	1.5%	2.0%	5.0%
Broken hip	1.0%	1.4%	0.3%	0.7%	0.9%	0.4%	-3.7%	-4.3%	8.0%
Parkinsons	1.6%	1.4%	1.1%	1.7%	1.5%	0.9%	0.6%	1.0%	0.2%
Emphysema/asthma/chronic obstructive pulmonary disease (COPD)	15.5%	19.5%	10.6%	19.5%	21.0%	13.3%	2.9%	1.0%	3.0%
Paralysis in past year	1.8%	1.7%	0.2%	1.8%	1.4%	0.0%	0.0%	-1.8%	2.6%
Mental retardation (excl. Alzheimers/dementia)	3.1%	0.9%	0.0%	3.0%	1.1%	0.1%	-0.2%	3.2%	40.7%
Renal failure	0.8%	4.0%	0.5%	1.0%	10.1%	1.1%	3.7%	12.4%	13.1%
Myocardial infarction/Heart attack	2.6%	2.1%	1.5%	2.0%	1.8%	6.0%	-2.3%	-1.4%	23.4%
Other heart conditions, valve problem	*5.6%*	*22.1%*	14.3%	4.3%	23.9%	7.2%	-3.0%	1.0%	-7.5%
Congestive heart failure	*3.3%*	*12.5%*	2.9%	3.5%	11.3%	2.5%	1.0%	-1.2%	-1.6%
Heart rhythm problem	*7.5%*	*19.3%*	6.2%	6.7%	21.1%	6.5%	-1.3%	1.2%	1.2%
Stroke/transient ischemic attack (TIA)	*3.5%*	*8.1%*	2.7%	2.5%	7.2%	4.5%	-4.1%	-1.4%	8.6%
Skin cancer	4.9%	5.0%	4.1%	5.7%	6.0%	7.3%	2.2%	2.3%	7.7%

Note: prevalences are weighted by respective survey weights.

Table 6: Prevalence in MEPS by year 2001-2009

Condition	2001	2002	2003	2004	2005	2006	2007	2008	2009
Hardening of arteries/arteriosclerotic heart disease + Angina/CHD	4.26%	4.49%	3.93%	4.35%	4.08%	3.80%	10.18%	18.18%	17.99%
Lung cancer	0.40%	0.51%	0.22%	0.49%	0.35%	0.40%	0.76%	0.78%	0.45%
Colon cancer	0.46%	0.43%	0.90%	0.60%	0.66%	0.72%	0.93%	0.92%	0.85%
Breast cancer	1.59%	1.84%	1.52%	1.72%	1.37%	1.23%	1.37%	2.19%	2.46%
Prostate cancer	1.80%	1.98%	1.52%	2.11%	1.80%	1.25%	1.79%	3.04%	2.97%
Other cancer	4.90%	4.85%	4.95%	5.23%	5.26%	5.08%	5.34%	6.43%	5.83%
Diabetes	15.95%	17.92%	17.19%	18.85%	19.99%	20.61%	22.27%	22.09%	21.83%
Arthritis	3.45%	3.26%	3.41%	2.95%	2.95%	3.50%	9.58%	16.26%	15.33%
Mental/psychiatric disorder (excl. Alzheimers/dementia)	13.70%	16.92%	17.22%	17.13%	17.98%	19.32%	19.56%	18.40%	18.88%
Alzheimers/dementia	1.79%	1.71%	1.23%	2.00%	2.44%	2.48%	1.95%	2.76%	2.46%
Osteoporosis	5.50%	7.26%	7.19%	6.28%	7.81%	7.76%	6.45%	6.55%	5.87%
Hypertension	43.01%	46.56%	47.94%	49.58%	52.19%	52.87%	57.24%	61.15%	63.24%
Broken hip	0.35%	0.68%	0.41%	0.43%	0.42%	0.46%	0.33%	0.51%	0.35%
Parkinsons	1.09%	1.11%	1.03%	0.92%	1.13%	0.79%	0.66%	0.63%	0.92%
Emphysema/asthma/chronic obstructive pulmonary disease (COPD)	10.62%	10.95%	9.60%	9.71%	10.73%	11.80%	12.16%	12.73%	13.26%
Paralysis in past year	0.21%	0.35%	0.15%	0.32%	0.18%	0.22%	0.30%	0.21%	0.04%
Comatose	0.04%	0.05%	0.00%	0.13%	0.25%	0.02%	0.00%	0.06%	0.01%
Mental retardation (excl. Alzheimers/dementia)	0.03%	0.02%	0.06%	0.10%	0.12%	0.26%	0.11%	0.08%	0.06%
Renal failure	0.54%	0.79%	0.68%	1.08%	1.14%	0.73%	0.65%	0.88%	1.07%
Myocardial infarction/Heart attack	1.48%	1.75%	1.80%	1.79%	1.79%	1.73%	3.54%	5.55%	6.04%
Other heart conditions, valve problem	14.34%	13.98%	12.77%	13.08%	12.12%	12.80%	9.49%	7.09%	7.20%
Congestive heart failure	2.90%	3.19%	3.16%	3.06%	3.12%	3.12%	2.77%	2.35%	2.48%
Heart rhythm problem	6.16%	7.31%	6.90%	7.02%	7.41%	7.79%	7.12%	6.12%	6.51%
Stroke/transient ischemic attack (TIA)	2.66%	3.10%	3.25%	2.99%	2.56%	2.48%	3.86%	4.77%	4.46%
Skin cancer	4.13%	4.45%	4.55%	5.17%	5.46%	5.70%	6.07%	7.75%	7.28%

Note: Prevalences are weighted by MEPS survey weights.

2009 with the treated prevalences in the MEPS for the same years and illnesses. The MEPS treated prevalences are calculated from the MEPS medical events files and a respondent is therefore only counted as having the condition if it was associated with a medical event.[2] The treated prevalences of some illnesses in the MEPS (such as heart disease, diabetes and arthritis) are much lower than the corresponding prevalences in the MCBS in 2001 but the two datasets have fairly similar prevalences for those illnesses in 2009. [3] Table 6 shows treated prevalence in the MEPS for every year from 2001 to 2009; the prevalences increase steadily with a sharp jump upward in 2007. By 2009, prevalences of these three illnesses have risen to levels comparable to those in the MCBS. This pattern may be explained by a change in the methodology for collecting diagnoses of chronic illnesses in 2007 that is discussed in the MEPS documentation.[4] However, as discussed above, any change in methodology for collecting diagnoses in a survey may have unfortunate implications for the correct measurement of an expenditure index from that data.

Method comparisons

As was shown in the previous two sections, we have a choice of methods and datasets available to us to calculate medical expenditure indexes for the Medicare population, although not every method will work with every dataset. We may also use a combination of methods and datasets for different subpopulations. Table 7 lays out the possible combinations of methods and datasets and shows the average annual growth rates of Fisher medical care expenditure indexes calculated from those methods and datasets for the years 2001-2009. The columns

[2] In the MEPS, conditions are also collected in the conditions file and total population prevalence can be calculated from that file. For more information on the two types of prevalences, see http://meps.ahrq.gov/mepsweb/survey_comp/MEPS_condition_data.pdf.
[3] Alzheimers/dementia, broken hip and paralysis also have much lower prevalences in the MEPS in 2001 but that may simply reflect the lack of nursing home residents in the MEPS as they do not have the same growth in prevalences between 2001 and 2009.
[4] See, for example, http://meps.ahrq.gov/mepsweb/survey_comp/survey_results_ques_sections.jsp?Section=PE&Year1=AllYear&Submit22=Submit.

Table 7: Average annual growth rates of Fisher price indexes 2001-2009 calculated on different datasets with different methods						
	MCBS				**MEPS**	
	FFS and HMO beneficiaries	**Full-year FFS beneficiaries**			**Full-year FFS beneficiaries**	**FFS and HMO beneficiaries**
	27 MCBS survey diagnoses	**27 MCBS survey diagnoses**	**27 MCBS claims diagnoses (CCS categories)**	**MCBS claims diagnoses (all CCS diagnoses)**	**MEPS Medicare beneficiaries (all CCS diagnoses)**	**MEPS Medicare beneficiaries (all CCS diagnoses)**
Regression-based with drug spending	2.9 (two-part) (A)	4.1 (two-part) (B)	3.0 (two-part) (C)	2.6 (two-part) (D)	2.4 (one-part) (F)	2.5 (one-part) (G)
Regression-based with drug spending, dropping MEPS events with no diagnosis	n/a	n/a	n/a	n/a	3.9 (H), 3.2 (H2-only counts dx if primary)	2.9 (I), 3.0 (I2-only counts dx if primary)
Regression-based without drug spending	2.3 (J)	3.4 (K)	2.2 (L)	1.9 (M), 2.5 (M2—only counts dx if primary)	3.2 (N)	3.3 (O)
Regression-based without drug spending, dropping MEPS events with no diagnosis	n/a	n/a	n/a	n/a	3.8 (V)	3.9 (W)
Primary diagnosis with drug spending	n/a	n/a	n/a	n/a	3.2 (P)	3.1 (Q)
Primary diagnosis without drug spending	n/a	n/a	2.5 (R)	2.8 (S)	4.3 (T)	4.2 (U)

represent different combinations of samples (FFS and private plan, FFS only), datasets (MCBS or MEPS) and sets of illnesses and the rows of the table represent different methods (regression-based or primary diagnosis, with or without drug spending). The illnesses used are either the 27 illnesses from the MCBS survey (see Hall and Highfill 2013 for a list and detailed discussion) or the 259 categories from the Clinical Classifications System (CCS), a system devised by the Agency for Healthcare Research and Quality for classifying the 10,000 or so ICD-9 diagnosis codes used in claims datasets into medical conditions. In one column we use 27 CCS diagnoses that correspond to the MCBS survey diagnoses; see Appendix Table 1 in Hall and Highfill (2013) for a crosswalk that translates survey diagnoses into claims diagnoses. For ease of comparison across years, when creating the FFS-only sample, we restricted the sample to beneficiaries enrolled in Medicare for the full calendar year. The downside of this approach is that we lose beneficiaries who pass away during the year and who are responsible for a good part of total Medicare spending.

The cells in the table are labeled with letters for ease of reference.

A: This index was presented previously in Hall and Highfill (2013). As explained in more detail there, it is calculated from the spending of both FFS and private plan beneficiaries with a regression-based method and a two-part model of health-care spending as a function of diagnoses. The total index is aggregated from indexes for 27 illnesses asked about in the medical conditions survey but prevalence information is supplemented with diagnosis codes from the attached medical claims, especially for beneficiaries who pass away during the year and who therefore do not have survey answers for that year.

B: This index is calculated with the same method as A but with a different sample: beneficiaries who are enrolled in FFS Medicare for the full year.

C: This index is calculated with the same sample as B and the same overall method as A but with the 27 illnesses defined from their equivalents in the ICD-9 codes in the claims data, not from the in-person survey. As discussed above, defining illnesses from the diagnosis codes in claims data instead of from in-person surveys can make a substantial difference to prevalence rates and to who is diagnosed with each illness. This issue will be discussed further below.

D: This index is calculated with the same sample as B and C but, instead of being an aggregation of indexes for only 27 illnesses, it is aggregated from indexes for the 259 categories of the CCS.

F: This index is calculated with the same method and illnesses as D, but on a sample from MEPS that is made to look as much like the sample in D from the MCBS as possible, as it only includes beneficiaries enrolled in FFS Medicare for all twelve months of the year. Note that one difference between the diagnoses used in B, C and D and those used in F is that diagnoses in the MEPS may also come from prescription drug events. The spending variables used in the regressions in B, C and D include drug spending (collected in the in-person survey part of the MCBS) but the diagnoses that underlie the dummy variables on the right-hand side of the regressions only come from doctor and hospital claims.

G: This index is calculated with the same method and illnesses as F but also includes Medicare private plan enrollees.

H, H2, I, I2: These indexes are calculated on the MEPS and have the same beneficiaries as the row above (F and G) but events with no diagnosis are not included in the spending totals on the left-hand side of the regression. H2 and I2 also only count a diagnosis on the right-hand side if it is a primary diagnosis. This is to make the index more comparable to cells P and Q which are calculated with the primary-diagnosis method which only includes the first diagnosis on an event with multiple diagnoses.

J, K, L, M, M2, N, O: These indexes are calculated with a regression-based method, as in the first row of the table, but drop drug spending, in order to directly compare their results with cells R, S, T, and U which also omit drug spending.

V, W: These indexes are the same as N and O except we drop events with no diagnosis in order to compare directly with T and U.

P, Q: These indexes are calculated with the primary diagnosis method on MEPS data.

R, S: These indexes are calculated on the Medicare Part A and Part B claims data from the MCBS with the primary diagnosis method. Since the claims do not include drug events or spending,

however, the index does not include drug spending either and is only calculated for purposes of comparison with T and U. Note these are the only indexes able to be calculated on the MCBS with the primary diagnosis method.

T, U: These indexes are the same as P and Q but drop drug spending and events in order to be compared directly with R and S.

The first step in our comparisons is to compare the price indexes obtained using the regression-based method with the primary diagnosis method when we use them on the same samples and the same illnesses. As noted above, the primary diagnosis method is probably preferable but can only be used when the data meet its stringent requirements. The main purpose of this comparison therefore is to see if the regression-based method gives similar or very different results from the primary diagnosis method and if it is an adequate substitute for the primary diagnosis method when the data do not have a diagnosis attached to each event or claim.

There are three pairs of analyses to compare, as summarized in Table 8 which shows their average annual growth rates from 2001-2009 as reported in Table 7:

Table 8: Average annual growth rates for selected medical expenditure indexes 2001-2009

Sample	Regression-based	Primary diagnosis
MEPS FFS beneficiaries	3.2 (H2)	3.2 (P)
MEPS FFS and private plan beneficiaries	3.0 (I2)	3.1 (Q)
Medicare claims (omits drug spending)	2.5 (M2)	2.8 (S)

As the table shows, both methods produce strikingly similar results in all three samples, especially after taking into consideration the probable size of the confidence intervals. Based on these results, it seems therefore that the regression-based method is an adequate substitute for the primary diagnosis method when calculating an aggregate price index.

However, another potential purpose of a medical expenditure index is to see how much the average expenditures and prevalences of individual illnesses are changing over time and here we see some differences between the two methods. Table 9 shows the correlations in price and per capita spending assigned to illnesses by the two methods, both in their level (averaged over 2001 and 2009) and in their growth rates. As the table shows, the correlations between the levels of prices and per capita spending are strongly positive but the correlations between the growth rates in prices and in per capita spending are close to zero and are sometimes negative. These relationships hold whether we are using the MCBS or the MEPS as our data source.

Another way to compare the methods is to compare which illnesses are assigned the most spending by each method and which illnesses contribute the most to the growth in per capita health-care spending from 2001 to 2009. Table 10 presents this comparison for the two indexes using the MEPS and including both FFS and private plan beneficiaries (cells I2 and Q in table 7). The left panel shows the illnesses that have the highest per capita spending in 2001 as given by both indexes; specifically, it shows the union of two sets of the ten illnesses with the highest per capita spending in 2001 from both indexes. The illnesses in bold are the overlap illnesses, those illnesses that are among the ten with the highest per capita spending in both indexes. The two methods both assign the highest per capita spending to hypertension and the second highest to diabetes, and in both, hypertension and diabetes are both far ahead of the third-ranked illness. They agree in placing four other illnesses among the ten with the highest per capita spending (other non-traumatic joint disorders, other and ill-defined heart disease, disorders of lipid metabolism, and congestive heart failure).

The right-hand panel shows the illnesses contributing the most to the growth in per capita spending. To explain what is meant by this, recall that total health-care spending is the average price of each illness times the number of people with each illness summed up over all illnesses.

$$\$ = \sum_{j=1}^{J} \bar{P}_j N_j$$

Table 9: Correlations between regression-based indexes and primary diagnosis indexes

Population	Percent correlation in price	Percent correlation in per capita spending	Percent correlation in price growth	Percent correlation in per capita spending growth
MEPS FFS beneficiaries (H2 and P)	80.8%	96.4%	8.1%	2.1%
MEPS FFS and HMO beneficiaries (I2 and Q)	77.4%	94.7%	-18.7%	-12.9%
MCBS FFS claims (M and S)	79.2%	62.6%	1.0%	2.5%

Notes: 1. Correlations are weighted by disease prevalence. 2. Correlations in levels (the first two columns) are averages of the 2001 correlation and the 2009 correlation.

Table 10: Leading illnesses in indexes calculated from the MEPS (FFS and HMO beneficiaries)

Top illnesses in 2001 (by per capita spending, shown)	I2	Q	Top illnesses contributing to growth in per capita spending 2001-2009 (percent contribution shown)	I2	Q
Essential hypertension	11.2%	6.0%	**Essential hypertension**	17.9%	8.5%
Diabetes mellitus without complication	6.4%	4.6%	**Disorders of lipid metabolism**	12.7%	6.6%
Other non-traumatic joint disorders	3.7%	3.6%	**Coronary atherosclerosis and other heart disease**	12.2%	11.8%
Residual codes, unclassified	3.6%	1.5%	**Diabetes mellitus without complication**	5.5%	3.5%
Other and ill-defined heart disease	3.4%	3.8%	**Affective disorders**	4.8%	3.9%
Disorders of lipid metabolism	3.3%	2.4%	Multiple sclerosis	3.2%	2.4%
Cataract	2.5%	1.8%	**Rheumatoid arthritis and related disease**	2.7%	3.0%
Other disorders of stomach and duodenum	2.0%	0.9%	Asthma	2.6%	0.8%
Administrative/social admission	1.9%	0.8%	Biliary tract disease	2.5%	1.3%
Congestive heart failure, nonhypertensive	1.9%	2.4%	Osteoarthritis	2.4%	1.5%
Cardiac dysrhythmias	1.4%	2.6%	Spondylosis, intervertebral disc disorders, other back	0.9%	3.9%
Pneumonia (except that caused by tuberculosis or sexually	1.6%	2.1%	Other non-traumatic joint disorders	1.3%	3.8%
Aortic and peripheral arterial embolism or thrombosis	1.1%	2.0%	Conduction disorders	0.5%	3.3%
Acute myocardial infarction	0.9%	2.0%	Acute and unspecified renal failure	-1.2%	3.1%

Per capita spending may therefore be written (where N is the total population):

$$\frac{\$}{N} = \sum_{j=1}^{J} \bar{P}_j \frac{N_j}{N}$$

$\frac{N_j}{N}$ is simply the prevalence, so per capita spending on each illness is therefore its price times its prevalence and total per capita spending may be understood as the sum of per capita spending on each illness.

The difference in total per capita spending between 2001 and 2009 may therefore be written:

$$\frac{\$_{09}}{N_{09}} - \frac{\$_{01}}{N_{01}} = \sum_{j=1}^{J} \bar{P}_{j09} \frac{N_{j09}}{N_{09}} - \sum_{j=1}^{J} \bar{P}_{j01} \frac{N_{j01}}{N_{01}}$$

$$= \sum_{j=1}^{J} (\bar{P}_{j09} \frac{N_{j09}}{N_{09}} - \bar{P}_{j01} \frac{N_{j01}}{N_{01}})$$

The difference in the sums is the sum of the differences and each illness has a contribution to the increase in total per capita spending which can be expressed as a percentage of that increase. Growth in per capita spending is therefore a function of growth in price and growth in prevalence.

According to both methods, hypertension is the illness that is the single biggest contributor to growth in per capita spending in the MEPS although the regression-based method assigns it a share of growth that is twice as large as that assigned by the primary diagnosis method (18% versus 9%). The two methods agree in placing five more medical conditions (disorders of lipid metabolism, coronary atherosclerosis, diabetes mellitus without complications, affective disorders, and rheumatoid arthritis) in the ten illnesses contributing the most to growth in total per capita spending.

Table 11 presents the comparison for the indexes that are calculated from MCBS FFS claims (M and S). Because the Medicare FFS claims do not include pharmaceutical claims, these indexes do not include any pharmaceutical spending or diagnoses from pharmaceutical claims or

Table 11: Leading illnesses in indexes calculated from MCBS FFS claims

Top illnesses in 2001 (by per capita spending, shown)	M2	S	Top illnesses contributing to growth in per capita spending 2001-2009 (percent contribution shown)	M2	S
Other lower respiratory disease	5.5%	1.3%	**Other connective tissue disease**	7.3%	2.6%
Nonspecific chest pain	4.4%	1.7%	**Residual codes, unclassified**	7.1%	2.7%
Coronary atherosclerosis and other heart disease	4.2%	6.3%	Deficiency and other anemia	6.7%	1.9%
Chronic renal failure	3.9%	4.0%	Medical examination/evaluation	5.4%	0.4%
Medical examination/evaluation	3.4%	0.2%	Other lower respiratory disease	4.9%	1.6%
Complication of device, implant or graft	2.6%	3.2%	Other aftercare	4.7%	0.8%
Chronic obstructive pulmonary disease and bronchiectasis	2.3%	1.9%	**Other nervous system disorders**	4.4%	4.1%
Cataract	2.3%	2.6%	Diabetes mellitus without complication	4.4%	1.0%
Essential hypertension	2.2%	1.2%	**Spondylosis, intervertebral disc disorders, other b**	4.3%	7.2%
Diabetes mellitus without complication	2.2%	0.8%	Urinary tract infections	3.1%	1.6%
Osteoarthritis	1.0%	2.8%	Spondylosis, intervertebral disc disorders, other back problems	4.3%	7.2%
Spondylosis, intervertebral disc disorders, other back problems	2.1%	2.5%	Chronic renal failure	-2.2%	5.4%
Cardiac dysrhythmias	2.2%	2.5%	Rehabilitation care, fitting of prostheses, and adjustment of devices	-2.6%	4.6%
Acute myocardial infarction	1.2%	2.3%	Septicemia (except in labor)	-0.4%	3.4%
Congestive heart failure, nonhypertensive	1.0%	2.2%	Osteoarthritis	1.7%	3.4%
Pneumonia (except that caused by tuberculosis or sexually transmitted disease)	1.4%	1.9%	Retinal detachments, defects, vascular occlusion, and retinopathy	2.9%	2.5%

events, so the set of illnesses that are leading in terms of per capita spending in 2001 or in contributing to growth are different from those in the indexes calculated from the MEPS data. In particular, hypertension and diabetes are much less important. As the table shows the two methods agree on four of the ten illnesses with the highest share of per capita spending in 2001 (coronary atherosclerosis; chronic renal failure; complication of device, implant or graft; and cataracts) and on four of the ten illnesses with the highest share of growth in per capita spending from 2001 to 2009 (other connective tissue disease; residual codes; other nervous system disorders; and spondylosis, intervertebral disc disorders, other back problems).

Since two major chronic illnesses, hypertension and diabetes, are assigned a larger share of spending both of the level and of the growth of per capita spending by the regression-based method than by the primary diagnosis method in the MEPS data, we explored whether the two methods generally differed in how they treated chronic and acute illnesses. Table 12 shows a comparison of expenditure indexes for chronic and acute illnesses on all three data samples as well as cells N and T, which calculate expenditure indexes for the MEPS FFS-only sample without drug spending. (Appendix Table A lists which conditions are included in the chronic and acute categories.) For the cells that include drug spending, it does appear that the regression-based method does give noticeably higher growth rates for the expenditure index for chronic illnesses. This pattern is not evident in the expenditure indexes for acute illnesses, so there may be differences between how the two methods handle chronic illnesses and this issue is worthy of further exploration.

In conclusion, it seems that the regression-based method performs very similarly to the primary diagnosis method in the aggregate and there is a moderate amount of agreement between the two methods in terms of which medical conditions contribute the most to both the level and the growth rate of per capita spending. In addition, while we presented one regression-based approach, there is a considerable amount of discretion in how this approach can be applied, both in the modeling of spending as a function of diagnoses and in how the coefficients are used to divide up individuals' health-care spending. One potential approach going forward therefore may be to adjust the regression-based approach until the results match the primary

Table 12: Expenditure indexes for subaggregates of chronic and acute conditions (comparing methods)

	Aggregate price index		Acute price index		Chronic price index	
Population	**Regression-based**	**Primary diagnosis**	**Regression-based**	**Primary diagnosis**	**Regression-based**	**Primary diagnosis**
MEPS FFS and HMO (I2, Q)	3.0%	3.1%	-0.1%	0.8%	3.2%	1.5%
MEPS FFS only (H2, P)	3.2%	3.2%	0.8%	1.1%	2.9%	1.4%
MEPS FFS only, no drugs (N, T)	3.2%	4.3%	2.3%	3.6%	0.1%	1.5%
MCBS FFS only, no drugs (M2, S)	2.5%	2.8%	6.6%	2.4%	0.6%	1.7%

diagnosis method within a certain level of tolerance in the MEPS data, and then apply that adjusted approach to the larger sample and larger targeted population of the MCBS. Another approach may be to combine the two methods: use the primary diagnosis method on the Medicare Part A and B claims in the MCBS but use a regression-based method on the drug spending from the survey portion of the MCBS.

Comparison of expenditure indexes calculated from different datasets

As Table 7 shows, we also calculated price indexes with the same methods and on comparable populations from the MCBS and the MEPS, for the purposes of comparing the results. In general, the MCBS or the Medicare claims are preferable datasets for analyzing the Medicare population to the MEPS since they have larger samples, also include the nursing home population, and the claims do not seem to have the same underreporting issues as the MEPS does. However, comparing the datasets offers a useful check on the MEPS which has been used on other populations in the medical-care expenditure index literature.

Table 13: Average annual growth rates for selected medical expenditure indexes 2001-2009

Method	MCBS	MEPS
Regression-based including drug spending	2.6 (D)	2.4 (F)
Regression-based omitting drug spending	1.9 (M)	3.2 (N)
Primary diagnosis omitting drug spending	2.8 (S)	4.3 (T)

Table 13 shows the pairs of cells from Table 7 that directly compare datasets with the same method and comparable populations. The population in both cases is FFS Medicare beneficiaries who are enrolled in Medicare for the full calendar year. In the MCBS sample, we use claims for diagnoses so as to be able to compare on an illness-by-illness basis with the MEPS, in which conditions are coded in ICD-9 codes. As the first row of Table 13 shows, when

drug spending is included and we use a regression-based method, we get nearly the same average annual growth rate for the aggregate expenditure index in both datasets. However, when drug spending is dropped, the expenditure indexes calculated from the MEPS have a higher growth rate by a percentage point or more.

Table 14 shows the correlations in prices, prevalences, and per capita spending across illnesses between the expenditure indexes calculated from the MCBS and the MEPS. The first three columns show the correlations in the levels of these variables averaged across 2001 and 2009; as they show, the correlations are strongly positive. The next three columns, however, show the correlations in the growth rates of these variables from 2001 and 2009 which are much lower and often close to zero.

Since pharmaceutical spending is a large component of spending on many chronic illnesses, the fact that we calculate higher price indexes in the MEPS than in the MCBS only when we drop pharmaceutical spending may result from acute and chronic illnesses being treated differently in the two datasets. Table 15 compares the average annual growth rates of expenditure indexes for subaggregates of chronic and acute illnesses from 2001 to 2009. As it shows, when drug spending is omitted, the expenditure index for acute illnesses calculated from the MEPS has a higher growth rate than when drug spending is included. However, we see no similar pattern in the expenditure indexes for chronic illnesses where omitting drug spending reduces the growth rate of the expenditure index calculated from the MEPS. The pattern we observe in the aggregate may therefore come from some other source or represent random variation.

Conclusion

We have compared medical care expenditure indexes calculated from different datasets and using different methods. Our belief going into this research was that the primary diagnosis method was the best method for dividing up health-care expenditure by disease but that the Medicare Current Beneficiary Survey was the best dataset for analyzing Medicare beneficiaries, as it has the widest coverage and the most information on them. However, the primary

Table 14: Correlations across illnesses between the MCBS and the MEPS

Method	Percent correlation in price	Percent correlation in prevalence	Percent correlation in per capita spending	Percent correlation in price growth	Percent correlation in prevalence	Percent correlation in per capita spending growth
Primary diagnosis without drug spending (S & T)	54.9%	89.0%	53.0%	10.5%	11.4%	0.6%
Regression-based (with drugs) (D & F)	38.4%	69.4%	81.3%	-0.3%	11.4%	-0.3%

Notes: 1. Correlations are weighted by disease prevalence. 2. Correlations in levels (the first three columns) are averages of the 2001 correlation and the 2009 correlation.

Table 15: Expenditure indexes for subaggregates of chronic and acute conditions (comparing datasets)

	Aggregate price index		Acute price index		Chronic price index	
Method	**MCBS**	**MEPS**	**MCBS**	**MEPS**	**MCBS**	**MEPS**
Regression-based (with drugs) (D & F)	2.6%	2.4%	4.5%	0.6%	3.8%	1.6%
Regression-based (without drugs) (M & N)	1.9%	3.2%	4.0%	2.3%	2.0%	0.1%
Primary diagnosis (without drugs) (S & T)	2.8%	4.3%	2.4%	3.6%	1.3%	1.5%

diagnosis method cannot be used with the MCBS, at least not with drug spending. We therefore compared the primary diagnosis method with a regression-based method for estimating expenditures by disease and found that, when they are run on the same datasets, they produce virtually identical results in the aggregate. On an illness-by-illness basis, there is a moderate level of agreement on which illnesses are the leading contributors to growth in per capita health-care spending. It seems therefore that a regression-based index is an acceptable substitute for the primary diagnosis method and we can use it on the MCBS to produce a medical care expenditure index.

We also compared medical care expenditure indexes for Medicare beneficiaries produced from the MCBS and the MEPS. Although, as noted above, the datasets differ in their coverage and the MEPS appears to have some problems with underreporting and methodology changes in collecting diagnoses during the time period we are studying, indexes produced by the same method from both datasets have very similar aggregate growth rates as long as drug spending is included. When drug spending is omitted, however, the MEPS produces an index with an average annual growth rate that is at least a percentage point higher than the index from the MCBS. Comparison of indexes calculated from subaggregates of acute and chronic illnesses shows that part of the increased growth of the MEPS index when drug spending is omitted may come from the MEPS treating chronic illnesses differently.

The best solution for FFS Medicare beneficiaries, in the end, may be a hybrid index: one that combines the primary diagnosis method applied to the Part A and Part B claims with a regression-based index for pharmaceutical spending. For private-plan beneficiaries, however, the solution is not as clear. For this population, our choices are between a regression-based method run on the MCBS with only the 27 diagnoses included in the MCBS survey, or the primary diagnosis or regression-based method run on the small group (about 300-400 a year) of Medicare private-plan enrollees in the MEPS. Exploring these options will probably be the subject of some of our future work.

While we believe this comparison has covered the most important methods and datasets, it is not comprehensive. In particular, we did not cover the use of commercial groupers and we did not explore the larger sample of Medicare claims that are available for research (including the Part D pharmaceutical claims). We hope to address these gaps in future work.

References

Aizcorbe, A., R. Bradley, R. Greenaway-McGrevy, B. Herauf, R. Kane, E. Liebman, S. Pack, and L. Rozental, "Alternative Price Indexes for Medical Care: Evidence from the MEPS Survey", BEA working paper, 2011.

Aizcorbe, A. and N. Nestoriak, "Changing mix of medical care services: Stylized facts and implications for price indexes", *Journal of Health Economics* 30:3 (2011): 568-574.

Congressional Budget Office, "The Budget and Economic Outlook: Fiscal Years 2013 to 2023", Feb. 13, 2013 (http://www.cbo.gov/publication/43861).

Dunn, A., E. Liebman, S. Pack, and A. Shapiro, "Medical Care Price Indexes for Patients with Employer-Provided Insurance: Nationally-Representative Estimates from MarketScan Data", BEA working paper, 2010.

Hall, A. and T. Highfill, "A Regression-Based Medical Care Expenditure Index for Medicare Beneficiaries", BEA working paper, 2013.

Trogdon, J., E. A. Finkelstein, and T. J. Hoerger, "Use of Econometric Models to Estimate Expenditure Shares", *Health Services Research* 43:4 (August 2008): 1442-1452

Appendix Table A

Illnesses in the acute subaggregate in Tables 12 and 15

Acute and unspecified renal failure
Pneumonia (except that caused by tuberculosis or sexually transmitted disease)
Acute cerebrovascular disease
Aortic, peripheral, and visceral artery aneurysms
Aortic and peripheral arterial embolism or thrombosis
Acute myocardial infarction
Sprains and strains
Fracture of lower limb
Acute bronchitis
Urinary tract infections
Respiratory failure, insufficiency, arrest (adult)
Fracture of neck of femur (hip)
Other injuries and conditions due to external causes
Allergic reactions
Viral infection

Illnesses in the chronic subaggregate in Tables 12 and 15

Essential hypertension
Disorders of lipid metabolism
Osteoarthritis
Diabetes mellitus without complication
Deficiency and other anemia
Malaise and fatigue
Chronic obstructive pulmonary disease and bronchiectasis
Thyroid disorders
Esophageal disorders
Chronic renal failure
Affective disorders
Other mental conditions
Multiple sclerosis
Rheumatoid arthritis and related disease
Anxiety, somatoform, dissociative
Hyperplasia of prostate
Glaucoma
HIV infection
Chronic ulcer of skin