

Predictability and predictiveness in health care spending

Randall P. Ellis^{a,*}, Thomas G. McGuire^b

^a *Economics Department, Boston University, 270 Bay State Road, Boston, MA 02215, USA*

^b *Harvard University, USA*

Received 12 January 2006; received in revised form 23 June 2006; accepted 30 June 2006

Available online 14 August 2006

Abstract

This paper re-examines the relation between the predictability of health care spending and incentives due to adverse selection. Within an explicit model of health plan decisions about service levels, we show that predictability (how well spending on certain services can be anticipated), predictiveness (how well the predicted levels of certain services contemporaneously co-vary with total health care spending), and demand responsiveness all matter for adverse selection incentives. The product of terms involving these three measures of predictability, predictiveness, and demand responsiveness define an empirical index of the direction and magnitude of selection incentives. We quantify the relative magnitude of adverse selection incentives bearing on various types of health care services in Medicare. Our results are consistent with other research on service-level selection. The index of incentives can readily be applied to data from other payers. © 2006 Elsevier B.V. All rights reserved.

JEL classification: I11; C21; D12

Keywords: Health plans; Adverse selection; Medicare; Managed care

1. Introduction

In the “open enrollment” policy environment characterizing Medicare, Medicaid, and private employers in the U.S., decisions of beneficiaries determine the distribution of health care cost risks among plans. Managed care plans paid by capitation have the means and the motive to influence these decisions, however, by rationing some services more tightly than others. For example, a plan may create a limited network of mental health care providers or other forms of specialty care

* Corresponding author. Tel.: +1 617 353 2741; fax: +1 617 353 4449.

E-mail address: ellisrp@bu.edu (R.P. Ellis).

in order to discourage enrollment among those who expect to use these services. Anticipating which services are likely to be subject to under (or over) provision could complement efforts to encourage efficient quality in managed care through quality reporting, through supplemental payments for achieving quality objectives, or through risk adjustment of plan payments.¹ This paper develops a way to classify services by deriving a simple service-level index measuring incentives to ration tightly and loosely. We implement this index using national Medicare data for 1996 and 1997.

Our index is related to how “predictable” are expenditures on a service, but as we show, predictability is only part of the story. Predictability is necessary for service-level rationing to work—if consumers cannot predict what they would use, they cannot be influenced in their choices by selective rationing—but predictability of a service only indicates what services a plan could usefully ration, not whether a plan would want to ration the service selectively, and if so, whether it would want to ration it tightly or loosely. These latter matters are governed by whether use of a service is “predictive” of total spending, that is, if persons using the service at issue are likely to use a great deal or not very much of all other services. As we will show, services that are both predictable and predictive of high total spending are in the most danger of being underprovided in managed care.

In Section 2, we briefly review of the literature in health economics that has focused until recently on “predictability.” Section 3 introduces a model of health plan choice of spending that shows how “predictability” works together with “predictiveness” and demand response to determine incentives for service-level adverse selection. Sections 4 and 5 describe the data and the empirical implementation of the index. Discussion of the implications for research on risk adjustment and other areas concludes the paper.

2. Studies of the predictability of health care spending

As is well known, the predictability of some of health care spending can interfere with the smooth functioning of insurance markets. In the literature in health economics, “predictability” typically refers to how much of the variation in individual level spending can be explained using predetermined person-level data. Early research on risk selection focused on what a health plan can predict about total health care costs, and how the plan might use this information to select enrollees. In an early theoretical paper, [Feldman and Dowd \(1982\)](#) assume that plans can accurately forecast potential enrollees’ costs, and can select only those enrollees for whom the expected costs are less than the premium the plan is paid. In their model, HMO enrollment is supply determined: all potential enrollees are assumed to seek enrollment, and the plan is able to choose the ones it accepts.² The mechanism used by plans to actually achieve this selection are unspecified, an unsatisfying omission in light of the contractual requirement of open enrollment. Because this literature focuses on the plan’s decision to exclude individuals, and not on the plan’s decision about the nature of its product, concerns about the predictability of individual services or the correlation of service costs with total costs do not emerge.

¹ It is well-recognized that health plans have an incentive to ration some services more tightly than others ([Luft and Miller, 1988](#); [Newhouse, 2002](#)). In their review of the evidence on “quality” in capitation-based managed care plans, one of the few clear-cut findings is that managed care plans tend to provide a higher quality of care for preventive services ([Miller and Luft, 2002](#)). [Cutler \(2003\)](#) recommends explicit payments for measured quality. [Glazer and McGuire \(2002\)](#) lay out a method for designing risk adjustment to counter service-level selection.

² More recent examples of papers along the same lines are [Newhouse et al. \(1989\)](#) and [Shen and Ellis \(2002a,b\)](#).

A related empirical literature has focused on predictability of total expenditures, ultimately for purposes of comparison with explained variance in a risk adjustment formula. Newhouse et al. (1989), van Vliet (1992), Welch (1985), and McCall and Wai (1983) all regard “what is predictable” as a statistical problem, where there is an absolute standard defined by a true model that uses all available information from year 1 to predict total spending in year 2. An agent, the individual, health plan or regulator, might know some of the predictor variables and therefore be in a position to act on them.

There are some important differences in how the maximum predictable proportion of the variance is defined in these papers. Newhouse et al. (1989) define the “maximum lower bound” predictable to be the variance predictable by the full set of time-invariant individual characteristics. It is a maximum because it uses all the information in time invariant characteristics. It is a lower bound because it disregards any information in the time varying characteristics that might be available for prediction. van Vliet (1992) allows for the possibility of time-varying characteristics by incorporating autoregressive components of spending in his predictions. Welch (1985) employs a related idea in supposing that expenditures for any person in a year are in the process of regressing towards a mean for a person of that type. Papers using various data sets and making use of these assumptions conclude that the “predictable” component of variance is about 15–25%.³ Some papers have explored the importance of this degree of predictability and simulated health plan incentives to engage in selection of profitable enrollees (van Vliet, 1992; Newhouse, 1997; Shen and Ellis, 2002a). They all find that risk adjustment reduces but does not eliminate the profitability of risk selection.

Glazer and McGuire (2000) initiated a third line of research emphasizing the importance of service-level product distortions as a risk selection strategy. In their framework, health plans choose service levels so as to selectively ration quantities of each type of service, and consumers choose plans in light of their expectations of what they would receive in the plan. After taking account of any risk adjustment, health plans have an incentive to over provide services that are disproportionately anticipated by profitable (healthy) people and undersupply services that are disproportionately anticipated by unprofitable (chronically ill) people. In service-level selection, it is plans’ actions seeking to influence who demands membership that are the source of adverse selection. It is recognized that plans must accept all applicants and must therefore effectuate selection by choice of the mix of “quality” of its services. Frank et al. (2000), hereinafter, FGM, operationalize the Glazer and McGuire framework by deriving expressions for the profit-maximizing level of rationing for each service by a health plan. Their expressions and discussion incorporate ideas of predictability and predictiveness. In this paper, we build on their analysis to isolate the contributions of service-level predictability and predictiveness to selection incentives.

Empirical research on “service-level selection” confirms that plans act on these incentives. Mello et al. (2002) find evidence that HMOs ration hospital care more strictly than physician visits in relation to regular Medicare. Cao (2002) demonstrates that service-level spending by Medicare enrollees in the FFS sector predicts whether a person chooses to switch into an HMO, providing evidence that Medicare HMOs appear to use service-level product distortion strategies to selectively attract enrollees. Cao and McGuire (2003) examine the pattern of risks remaining in regular Medicare in markets with higher and lower managed care penetration and find that

³ This framework gives a more hopeful view of existing risk adjustment formula than one which simply compares variance explained by risk adjustment formula to total variance. For years Medicare paid on the basis of age, gender, county and Medicaid status, which together explain just one percent of variance in total spending.

the HMOs deter higher risk beneficiaries with respect to some services but not others. Spending patterns in the FFS sector seem to reflect service-level distortion, with Part A (hospital) spending on mental disorders, and Part B spending on psychiatry and general surgery showing the most significant impact of this service selection.

3. Health plan profitability, service predictability, and predictiveness

We now present a model of profit-maximization by a health plan to derive explicit expressions for the contributions of predictability and predictiveness to incentives to distort service-level quality. Assume there are N individuals choosing a health plan. In this section, we analyze the behavior of one (representative) health plan, taking the behavior of the other plans as given. The model presented in this section is based on FGM. The health plan is paid a premium (possibly risk-adjusted) for each individual that it enrolls. Individuals differ in their need/demand for health care, and choose a plan which maximizes their expected utility. “Health care” is not a single commodity, but a set of services-maternity, mental health, emergency care, cardiac care, and so on. A health plan chooses a rationing or allocation rule for each service. The plan’s choice of rules will affect which individuals find the plan attractive and will therefore determine the plan’s revenue and costs. We assume that the plan must accept every applicant. We are interested in characterizing the plan’s incentives to ration services and how this relates to the predictability and predictiveness of different services.

3.1. Utility and plan choice

The health plan offers s services. Individuals’ expectations about the value of services they would receive determine their choice of plan. Let \hat{m}_{is} denote the amount that individual i expects the plan will spend on providing service s to him, if he joins the plan, and let $\hat{m}_i = \{\hat{m}_{i1}, \hat{m}_{i2}, \dots, \hat{m}_{is}\}$. The dollar value of the benefits individual i gets from a plan, $u_i(\hat{m}_i)$, is composed of two parts, a valuation of the services an individual expects to get from the plan, and a component of valuation that is independent of services. We assume these enter additively in utility. Thus,

$$u_i(\hat{m}_i) = v_i(\hat{m}_i) + \mu_i \quad (1)$$

where

$$v_i(\hat{m}_i) = \sum_s v_{is}(\hat{m}_{is})$$

is the service-related part of the valuation and is itself composed of the sum of the individual’s valuations of all services offered by the plan. $v_{is}(\cdot)$ is the individual’s valuation of spending on service s , also measured in dollars, where $v'_{is} > 0$, $v''_{is} < 0$. The $v_{is}(\cdot)$ is also an expectation. The individual does not know with certainty how much he will benefit from care of various types.⁴ The non-service component is μ_i , an individual-specific factor (e.g., distance or convenience) affecting individual i ’s valuation, known to person i . From the point of view of the plan, μ_i is unknown, but is drawn from a distribution $\Phi_i(\mu_i)$. We assume that any premium the person pays has been predetermined and is not part of the strategy the plan uses to influence selection.

⁴ One simple interpretation along these lines is to write $v_{is} = \hat{p}_{is} v_s(\cdot)$. \hat{p}_{is} is subjective probability each individual i holds about needing service s , and $v_s(\cdot)$ is the benefits a person receives from services if they are ill and need service s .

The plan will be chosen by individual i if $\mu_i > \bar{\mu}_i$, where $\bar{\mu}_i$ is the valuation the individual places on the next preferred plan. We analyze the behavior of a plan that regards the behavior of all other plans as given, so that $\bar{\mu}_i$ can be regarded as fixed. Individual i chooses the plan if:

$$\mu_i > \bar{\mu}_i - v_i(\hat{m}_i).$$

For now, we assume that, for each i , the plan has exactly the same information as individual i regarding the individual's service-related valuation of its services, v_i , and regarding the utility from the next preferred plan, $\bar{\mu}_i$. The individual's expectations about services to be received in the plan can be assumed to be the same as the plan's beliefs about those expectations. For each individual i , the plan does not know the true value of μ_i but it knows the distribution from which it is drawn. Therefore, for a given \hat{m}_i and $\bar{\mu}_i$, the probability that individual i chooses the plan, from the point of view of the plan, is⁵:

$$n_i(\hat{m}_i) = 1 - \Phi_i(\bar{\mu}_i - v_i(\hat{m}_i)) \quad (2)$$

3.2. Managed care

Managed care rations the amount of health care a patient receives. Following Keeler et al. (1998) and FGM, let q_s be the service-specific shadow price the plan sets determining access to care for service s . A patient with a benefit function for service s of $v_{is}(\cdot)$ will receive a quantity of services, m_{is} , determined by:

$$v'_{is}(m_{is}) = q_s \quad (3)$$

Let the amount of spending determined by the equation above be denoted by $m_{is}(q_s)$. In a first best $q_s = 1$ and the marginal benefit of spending would equal its marginal cost. m_{is} is actual spending on service s the plan provides to person i , as distinct from \hat{m}_{is} , the spending the individual expects.

3.3. Profit and the effect of rationing on profit

Let $q = \{q_1, q_2, \dots, q_s\}$ be a vector of shadow prices the plan chooses, $m_i q = \{m_{i1}(q_1), m_{i2}(q_2), \dots, m_{is}(q_s)\}$ be the vector of spending individual i gets by joining the plan and $\hat{m}_i(q) = \{\hat{m}_{i1}(q), \hat{m}_{i2}(q), \dots, \hat{m}_{is}(q)\}$ be the vector the individual expects to get by joining the plan. Expected profit, $\pi(q)$, to the plan will depend on the individuals the plan expects to be members, the revenue the plan gets for enrolling these people, and the costs of each member.

$$\pi(q) = \sum_i n_i(\hat{m}_i(q)) \left[r - \sum_s m_{is}(q_s) \right] \quad (4)$$

where r is the revenue the plan receives for each individual. To focus on the characteristics of services that determine incentives, we assume there is no risk adjustment and the same premium is paid for every enrollee. Note that it is individuals' expectations, the

⁵ An alternative interpretation is that index i describes a group of people with the same $v_i(\hat{m}_i)$ function and $n_i(\hat{m}_i)$ is then the share of this group that joins the plan.

\hat{m}_i 's that determine enrollment, and plan's expectations about actual spending that determine expected costs. Since (4) is an expression for the plan's expected profits, as we noted above, the individuals' expectations are properly regarded as the plan's beliefs about individuals' expectations.

The plan chooses a vector of shadow prices to maximize expected profit, (4). Defining $M_i = \sum_s m_{is}(q_s)$ to be total spending on a person, profit per person is $\pi_i = r - M_i$. Define η_s to be the elasticity of demand for service s with respect to a change in q_s . All individuals are assumed to share the same elasticity of demand for any service but elasticities can differ across services.⁶ The derivative of profit with respect to q_s , $s = 1, 2, \dots, s$, is (see FGM):

$$\frac{\partial \pi}{\partial q_s} = \eta_s q_s \sum_i \phi_i \hat{m}_{is} \pi_i - \eta_s \sum_i n_i m_{is} \quad (5)$$

In order to simplify (5), normalize the number of enrollees in the plan at 1, and then $\sum_i n_i m_{is} = \bar{m}_s$, where \bar{m}_s is the average spending on service s in the plan. Following FGM we assume that the enrollment function is uniform and the same for all i so $\Phi'_i = \phi$. Also, we wish to characterize incentives to ration care in relation to some baseline q , so we evaluate (5) for each s at the first best $q_s = 1$. With these assumptions,

$$\frac{\partial \pi}{\partial q_s} = \bar{m}_s \eta_s \left(\phi \sum_i \frac{\hat{m}_{is} \pi_i}{\bar{m}_s} - 1 \right) \quad (6)$$

3.4. An index of incentives to raise the shadow price for a service

We now use this expression to define an index I_s that quantifies the profit incentive to raise q_s (ration more tightly) for service s .

$$I_s \equiv \frac{\partial \pi}{\partial q_s} \cdot \frac{1}{\bar{m}_s} \quad (7)$$

I_s characterizes the profit incentives to ration service s . It is natural to deflate this derivative by the mean level of expected spending on service s , since rationally services with larger spending will affect total profits only because they are large. Dividing by the mean spending converts the measure to per-dollar terms. Substituting (6) into (7) we have

$$I_s = \eta_s \left(\phi \sum_i \frac{\hat{m}_{is} \pi_i}{\bar{m}_s} - 1 \right) \quad (8)$$

Letting σ_x be the standard deviation of x , we write the correlation between \hat{m}_{is} and π_i as:

$$\rho_{\hat{m}_s, \pi} = \left(\frac{\sum_i \hat{m}_{is} \pi_i - \bar{m}_s \bar{\pi}}{\sigma_{\hat{m}_s} \sigma_{\pi}} \right) \quad (9)$$

⁶ Note that the fact that all individuals share the same elasticity of demand for a certain service does not imply that their demand curves are identical.

Using (9) Eq. (8) becomes

$$\begin{aligned} I_s &= \eta_s \left(\phi \frac{\sigma_{\hat{m}_s} \rho_{\hat{m}_s, \pi} + \bar{m}_s \bar{\pi}}{\bar{m}_s} - 1 \right) = \eta_s \sigma_{\pi} \phi \left(\frac{\sigma_{\hat{m}_s}}{\bar{m}_s} \rho_{\hat{m}_s, \pi} + \bar{\pi} - \frac{1}{\phi \sigma_{\pi}} \right) \\ &= \sigma_{\pi} \phi \eta_s \left(\frac{\sigma_{\hat{m}_s}}{\bar{m}_s} \rho_{\hat{m}_s, \pi} - C \right) \end{aligned} \quad (10)$$

where Eq. (10) implicitly defines C to capture terms that do not depend on s . Also note that $\sigma_{\pi} \phi$ is constant across s . With no risk adjustment, profit for each i is a constant less total spending, M_i , and $\rho_{\hat{m}_s, \pi} = -\rho_{\hat{m}_s, M}$. Thus, we can write our renormalized index of selectivity for service s as:

$$\tilde{I}_s = -\eta_s \left(\frac{\sigma_{\hat{m}_s}}{\bar{m}_s} \rho_{\hat{m}_s, M} - C \right) \quad (11)$$

3.5. Predictability, predictiveness and incentives to ration services

The index \tilde{I}_s in (11) is a unitless measure, the product of the demand elasticity and a term that includes coefficient of variation and a correlation coefficient.⁷ The coefficient of variation is of the individual's expected spending on service s , $\sigma_{\hat{m}_s}/\bar{m}_s$, and the correlation is between this expected spending with total spending, $\rho_{\hat{m}_s, M}$.

The coefficient of variation of the expected spending on service s , cv_s , represents the *predictability* of service s . The lower limit of predictability is when individuals do not expect to be different from the average, in which case $\sigma_{\hat{m}_s}$ is zero and the cv_s is zero. As individuals' ability to forecast their difference from the average increases, $\sigma_{\hat{m}_s}$ and cv_s go up. Later, we compute cv_s for each service by calculating the cv of the predicted spending on each service from regression models of spending.

The correlation between individuals' expected spending on a service s and total spending captures the *predictiveness* of service s , a straightforward interpretation. Note that the correlation could be positive or negative. Throughout this discussion, it is important to keep in mind that it is the plan's expectations about spending on services for individuals and the plan's beliefs about individual's expectations about what they will get upon joining that determine incentives affecting plans' behavior.

How predictability and predictiveness work together to determine incentives can be depicted in Fig. 1, where predictiveness of service s is depicted on the vertical and predictability of service s on the horizontal axis. The rectangular hyperbola where the product $\rho_{\hat{m}_s, M}(\sigma_{\hat{m}_s}/\bar{m}) = C$ characterizes a boundary of combinations of predictability and predictiveness of service where the incentives to ration services changes from oversupplying to undersupplying. The efficient level of rationing can be defined in principle at $q=1$ for all services, which would require a certain budget for the managed care plan to cover costs at this level of rationing. In our empirical work, we will instead use the standard of the average level of rationing. Whatever the budget a plan has to work with, a second-best efficient outcome is when the rationing is the same for all services.⁸ When we use the terms under or over provide we will mean them in the sense of comparing to this second-best standard.

⁷ The coefficient of variation is the standard deviation of a random variable divided by its mean.

⁸ This follows from maximizing the health benefits of a fixed group of enrollees subject to a budget for all services. See FGM (2000) for more discussion.

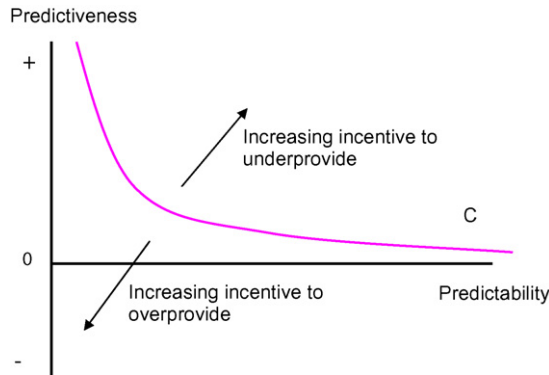


Fig. 1. Predictability and predictiveness determine the incentive to under(over)provide services.

When the product of the two terms for a service is greater than some constant C , the profit with respect to raising q_s is greater than for the average service. Conversely, when the product of predictability and predictiveness for a service is less than C , the profit gains from raising q_s are lower and the incentive to raise q_s is less. Lowering q in relation to the average is rationing loosely (in relation to the average.) As Fig. 1 makes clear, when a service is negative in terms of predictiveness, it will certainly be rationed loosely, but this is not a necessary condition for looser-than-average rationing.

The index in (11) also includes the term η_s , the elasticity of demand for service s . The demand elasticity does not affect the direction of incentives to over or underprovide, but does affect multiplicatively the magnitude of the incentives.

We now turn to data from Medicare that can be used to construct some elements of the service-level index (11). Predictability and predictiveness can be estimated from data. After doing so, we will discuss the effect different demand elasticities, which we do not estimate, would have on our interpretation of service level incentives to ration.

4. Data

We use the 1996 and 1997 Medicare Standard Analytical Files, which are a random 5% sample of all Medicare beneficiaries. From the Denominator files for 1996 and 1997 we select a sample of aged and disabled enrollees who were eligible for both Part A and Part B coverage for all of 1996 and at least part of 1997.⁹ Similar sample selection criteria have been used for risk adjustment calibration studies using Medicare populations, such as Ellis et al. (1996) and Pope et al. (2000,

⁹ Specifically we selected the sample with the following characteristics:

- Not eligible for coverage as part of the End Stage Renal Disease program
- Eligible for both Part A and Part B coverage for all 12 months of 1996
- Eligible for both Part A and Part B coverage for at least 1 month in 1997
- Age calculated from date of birth of less than 120
- Valid values of gender (male/female)

These selection criteria ensure that there is a full year of Part A and Part B utilization information for 1996, and some coverage to be predicted in 1997.

2004).¹⁰ We use information from the Denominator file to create age, gender, months eligible in 1996 and 1997, and Medicaid status.

For each individual in our sample we extracted claims information for 1996 and 1997 from the hospital discharges (Medpar), outpatient facility, home health care (Part A and Part B), skilled nursing facilities, and provider services (Part B). For this project we focus on total covered costs rather than Medicare payments, as Pope et al. (2000) and others have done. Our rationale is that we want to measure resources used in treating patients rather than measure what Medicare chooses to pay for. For Part B and for most Part A services the covered charge field is the amount that Medicare would cover before considering deductibles, copayments, coverage ceilings, or coordination of benefits with other insurers.¹¹

An important empirical question is how to divide spending into separate categories of “services” to evaluate selection incentives. Conceptually, one would like to partition spending along the lines of dimensions that health plans actually use to influence service provision. Health plans choose fees or discounts for each procedure, which suggests that groups of procedures, such as those classified by type of service, may be relevant. Health plans also contract selectively with different providers, and can potentially use provider specialty or place of service for selection. Finally, plans may be able to selectively manage the care of enrollees according to patient diagnoses, in which case spending by diagnosis may be relevant. We can see arguments for using each of these partitions: spending by type of service, spending by provider specialty, spending by place of service, and spending by diagnosis. Therefore, we explored each of these four dimensions and assessed their predictiveness and predictability.

To partition spending by type of service, we classified the procedure codes on each Part B claim using the Berenson–Eggers Type of Service (BETOS) codes, and then aggregated these BETOS values into 25 categories. Part A spending was aggregated into seven types of service values (room and board charges, hospital outpatient visits, home health care, hospice etc.), creating a total of 32 types of service (TOS) categories of spending.

In addition to partitioning spending by type of service, we also considered three other partitions of total spending. Our second partition grouped Part B spending according to provider specialty codes, using 29 mutually exclusive categories. Our third partition aggregated Part B spending by 39 places of service. Our final partition grouped spending by diagnostic categories. Unlike the other three breakdowns, dollars cannot be uniquely assigned to diagnoses, since many claims have multiple diagnoses on them, and some have none. We assigned spending on each claim to the first-listed diagnosis, which is often called the primary diagnosis. On hospital discharge records, the first-listed diagnosis has a special significance as the principal diagnosis. Diagnoses are missing on many claims, particularly on Part B claims, hence all such claims were aggregated into the residual category “invalid or no diagnosis recorded.” Spending by diagnosis was aggregated into condition categories using the Diagnostic Cost Group Condition Categories (CCs), which assigns each diagnosis to one of 183 condition categories, as described in Pope et al. (2000). We used DxCG software release 6.01 for our classification system.

¹⁰ However, Pope et al. (2000) excluded people in hospices and working aged beneficiaries, exclusions not made here.

¹¹ For Medicare inpatient facility charges, which are paid primarily using Diagnosis Related Groups (DRGs), there is no variable equivalent to covered charges, so we used total charges adjusted by the ratio of aggregate payment to aggregate charges. This adjustment better brings total hospital charges in line with the average Medicare payments, as others have done more precisely using hospital cost to charge ratios.

5. Empirical results

5.1. Methodology

Our analytical model develops an interpretation of the product of the coefficient of variation of predicted spending on a service and also the contemporaneous correlation of predicted spending on that service with total spending which we now estimate empirically. There is nothing in our theoretical framework that restricts how predictions are made by consumers or health plans, and therefore we contrasted five estimation methodologies. Our first and preferred method uses weighted least squares, where the dependent variable is annualized spending (actual spending divided by the fraction of the year eligible), and individuals with fractional years of eligibility are included and weighted by the fraction of the year eligible. This is the methodology described in [Ellis and Ash \(1995\)](#), and adopted by the US Medicare program.

Our second estimation approach uses simple OLS, and excludes people with fractional years of eligibility. This approach focuses on people with the most complete information, and uses a sample that is also used by the remaining three approaches. Our third estimation approach is a heteroskedasticity-corrected square root model as described in [Veazie et al. \(2003\)](#). The square root of actual covered charges is regressed on the given set of independent variables, and which generates predictions of the square root, which we call $G_i = X_i B$. Squared residuals from this regression are then regressed on the fitted value G . Heteroskedasticity-corrected predictions of spending from this model are the squared fitted predictions, G_i^2 , plus the predicted variance for each observation $s^2(G_i)$. Our fourth model uses a “two part OLS with smearing” as described in [Manning \(1998\)](#), and [Buntin and Zaslovsky \(2004\)](#). The first stage is a probit model of the probability of any spending, and the second stage is a linear model of spending among those with positive spending. Our fifth specification uses a generalized linear model (GLM) evaluated by [Buntin and Zaslovsky \(2004\)](#) with the linear portion transformed using the log transformation, and additive errors assumed to be normally distributed, hence $Y = \exp(X\beta) + \varepsilon$.¹²

Our estimation strategy is as follows. For each information set and estimation methodology, we first evaluate the how well the information and specification predicts total year two spending. For this we focus on the R^2 , as the most widely used and easily interpreted measure of predictability. We then choose two of these sets of information and four estimation methodologies to predict spending by service, calculating both the coefficient of variation of the predictions (predictability) and correlations with year two spending (predictiveness). We then calculate our selection indices and consider their rankings.

5.2. Predictive power of different information sets for total spending

It is plan beliefs and plan beliefs about consumer predictions that drive plans' decisions about risk selection. Unfortunately, we do not know what information consumers and plans actually use for their predictions, or whether linear or nonlinear predictive models are used. Therefore, we explored various possible information sets and econometric specifications to assess whether these choices are important.

¹² We also tried estimating two part log models and GLM models with alternative link and distribution functions however these models either did not converge or had very poor or negative R^2 values. Similar problems with these models were found in [Veazie et al. \(2003\)](#).

Table 1
Predictive power of various information sets and various models

	Weighted OLS	OLS	Square root model (hetero-skedastidty- corrected)	Two part linear model	GLM with link = log, dist = normal
Partial year eligibles included?	Yes	No	No	No	No
Sample mean	6886	5063	5063	5063	5063
Number of observations	1,380,863	1,273,471	1,273,471	1,273,471	1,273,471
	R^2				
Age and gender only	0.011	0.010	0.009	0.010	0.010
Prior year total covered charges ^a	0.089	0.096	0.113	0.120	0.105
Diagnoses organized by DCG/HCC ^a	0.104	0.108	0.103	0.107	0.105
Covered charges by DCG/HCC ^a	0.099	0.107	0.103	0.105	0.095
Covered charges by place of service ^a	0.140	0.145	0.136	0.145	0.126
Covered charges by physician specialty ^a	0.142	0.152	0.143	0.152	0.131
Covered charges by type of service ^a	0.150	0.155	0.146	0.154	0.134
All of the above except diagnoses ^a	0.154	0.160	0.151	0.160	0.138
“Kitchen sink”: all of the above ^a	0.169	0.171	0.161	0.169	0.147

Dependent variable: 1997 annualized total covered charges.

^a All regressions included a constant and 21 age-gender dummy variables.

Table 1 summarizes the predictability of various combinations of information sets and five estimation methodologies. Because they are widely used, we first estimated three types of risk adjustment models: an age and gender only model, a model using one total prior year spending variable, and an all-encounter diagnosis risk adjustment model using Diagnostic Cost Groups, as organized by Hierarchical Condition Categories (HCCs).¹³ Consistent with what others have found using Medicare samples (e.g., Pope et al., 2000) age and gender alone explain about 1% of the variability across individuals in the US Medicare population, regardless of the estimation method. Using age and gender with a single variable of lagged total covered charges alone explains about 9% of the total variation in 1997 annualized spending. Using age and gender together with 183 diagnostic dummy variables achieves a comparable level of predictiveness, explaining 10.4 to percent of the variation in total covered charges in our sample.¹⁴

Using disaggregated spending as the information set significantly improves the predictability of total health care spending. Among the four ways of decomposing total spending, using the previous year spending broken down spending by diseases (DCG/CCs) has the lowest predictive power ($R^2 = 10\text{--}11\%$) while spending by type of service has the highest ($R^2 = 13\text{--}15\%$).¹⁵ The bottom

¹³ All models include 22 age and gender categories.

¹⁴ For comparison, Pope et al. (2000) achieves an R^2 using data from the same 2-year period of .112, but uses payments rather than covered charges as the dependent variable.

¹⁵ This is probably related to the relatively large amount of spending with no diagnosis assigned, as well as the assignment of spending to only the first diagnosis listed on each claim.

two rows of Table 1 show that there is considerable overlap in the information contained in these different spending decompositions. The R^2 when all four sets of spending variables are included in the regression model only increases modestly to 15–16%, and when diagnostic information is added in, the R^2 rises to 16–17.¹⁶ It is reassuring that the five econometric specifications gives nearly the identical rankings in each case for the different information sets.

We conclude from this regression analysis of aggregate spending that there is considerable predictive power in using disaggregated rather than aggregated spending variables. Spending by TOS is marginally more predictive of subsequent year spending than other partitions of total spending, but all of the partitions have similar predictiveness. This finding is highly robust to different statistical specifications. There is relatively little to be gained by using one disaggregate over another. In separate analysis (not shown here) we found that the most predictive information set for each partition of total spending is generally the lagged spending broken down by that same partition. Hence lagged disaggregated spending by TOS is usually but not always more predictive of spending by TOS than lagged spending broken down by PS. Similarly, lagged spending by PS is generally more predictive of spending by PS than spending by TOS. We therefore used lagged disaggregated spending by TOS to predict TOS spending and lagged PS spending to predict PS spending. We also focus our analysis on the weighted least squares results that include partial year eligibles, although we return to the full year eligibility sample below.

5.3. Results by type of service (TOS)

In order to implement our proposed index, we next examined the predictability and predictiveness of spending by TOS categories, based on the Berenson/Eggers classification discussed earlier.¹⁷ Each row in Table 2 corresponds to spending on a distinct TOS. Weighted sample means are shown in column 1. Column 2 shows the conventional R^2 , calculated from weighted least squares regressions for each TOS using the full array of 32 lagged TOS spending variables together with age and gender dummies as explanatory variables. The third column shows our new measure of predictability: the coefficient of variation of predicted (not actual) spending on each TOS. Hospice spending and Part A inpatient visits have the highest predictability measures. Our predictability measure produces a different ranking than R^2 (correlation coefficient between these two measures is $\rho=0.28$). Some services, such as lab tests, intermediate care, and oncology services, have relatively low R^2 values (little of the variance across people is predictable) while they still have relatively high coefficients of variation (the explained variation is large relative to the mean). Our theoretical analysis suggests that the latter is more important.

The fourth column in the table displays our measure of predictiveness: the contemporaneous Pearson's correlation between predicted spending on the service and total spending. At this level of aggregation, the correlations are all positive, although with a different, finer breakdown, some would likely be negative. Some services that are highly predictable (hospice care and Part A inpatient visits) are not particularly predictive, while other services that are highly predictive (inpatient facility charges, including room and board) are not very predictable.

¹⁶ These last two results in particular overstate the R^2 due to overfitting and would not be validated using split sample techniques.

¹⁷ We dropped TOS and PS variables with less than \$4 per capita as well as the "other service" category. We also dropped dialysis since it is largely covered under a separate program and hence spending on this type of service is incomplete.

Table 2

Predictability and predictiveness of spending by type of service: weighted least squares results including partial year eligibles ($N=1,380,863$)

Spending by type of service categories	Mean (1)	Predictability		Predictiveness (4) $\text{Corr}(\hat{m}_s, M)$	Selection index (5) $(3) \times (4)$
		(2) R^2	(3) $CV(\hat{m}_s)$		
Hospice	4.56	0.554	53.879	0.048	2.578
Home health care	525.13	0.451	3.902	0.224	0.875
Durable medical equipment	194.65	0.474	4.007	0.175	0.703
Inpatient visits	47.16	0.260	6.540	0.091	0.592
Other (including unclassified)	112.15	0.274	4.885	0.101	0.495
Intermediate care	14.66	0.013	4.449	0.092	0.408
Hospital visit	126.98	0.074	1.121	0.318	0.356
Home visit	25.10	0.274	2.680	0.130	0.348
Inpatient facility (R&B)	2975.11	0.066	0.828	0.323	0.267
<i>Median service by type of service</i>					
Emergency room	26.45	0.153	1.167	0.227	0.265
Consultation	56.03	0.076	0.780	0.280	0.219
Other facility services	83.78	0.012	0.693	0.248	0.172
Outpatient visit	504.65	0.084	0.994	0.171	0.170
Advanced imaging CAT	15.81	0.067	1.239	0.136	0.169
Oncology	21.69	0.018	1.763	0.090	0.159
Lab tests	69.93	0.250	0.959	0.151	0.144
Other tests	33.52	0.137	1.053	0.128	0.134
Standard imaging	45.64	0.084	0.674	0.176	0.119
Specialist	57.84	0.227	1.615	0.071	0.114
Echography	33.46	0.097	0.882	0.128	0.113
Ambulatory procedures	35.81	0.032	0.831	0.127	0.105
Imaging procedure	7.94	0.017	0.864	0.118	0.102
Office visit	153.31	0.303	0.743	0.129	0.096
Major proc cardiovascular	54.81	0.015	0.844	0.113	0.096
Minor procedures	57.87	0.128	0.999	0.095	0.095
Anesthesia	34.31	0.024	0.512	0.180	0.092
Endoscopy	45.19	0.039	0.733	0.119	0.087
Major procedure	38.63	0.008	0.562	0.155	0.087
Major proc orthopedic	32.31	0.009	0.673	0.128	0.086
Advanced imaging – MRI	14.25	0.027	1.148	0.072	0.083
Eye procedures	75.40	0.018	0.619	0.072	0.045

Sample contains only individuals eligible for Medicare Part A and Part B for all of 1996 and for any part of 1997. Regression uses fraction of year eligible as sample weights. Type of service categories are aggregates of the Berenson–Eggers values.

The final column in the table shows our selection index, the product of our predictability and predictiveness measures for each service. To highlight the services that are most and least vulnerable to selection distortions, we have sorted services in this table and the next by their selection index. The results by TOS imply that spending on hospice care, home health care, durable medical equipment and inpatient visits are most vulnerable to underprovision. Spending on eye procedures, magnetic resonance imaging (MRI) major orthopedic procedures, major procedures, and endoscopy have the lowest selection indices, and hence the largest incentive for overprovision. These rankings have face validity to us: the overprovided services are those used primarily on relatively healthy people (often discontinued once they become ill), while the underprovided services are recognizable as those used primarily by high health cost people. Hospice care is

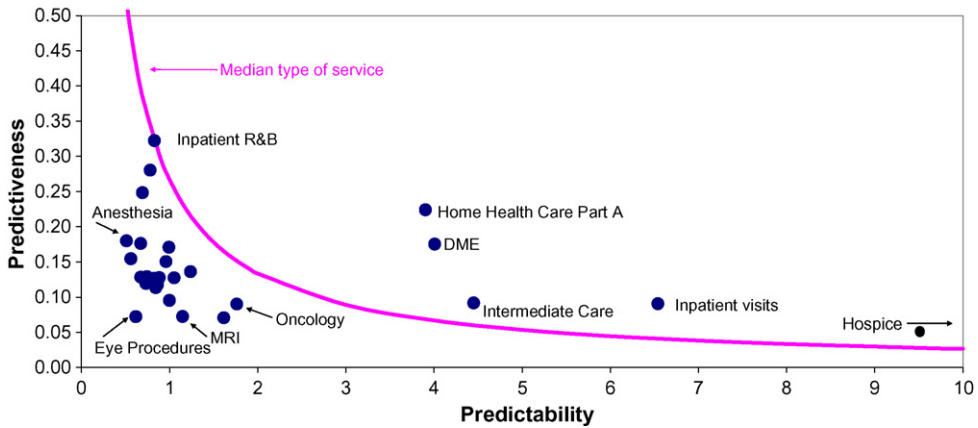


Fig. 2. Plot of predictability vs. predictiveness of spending by type of service.

carved out of Medicare health plans, and our analysis shows why this is necessary to avoid underprovision of that service.

The information in columns 3 and 4 of [Table 2](#) is also presented graphically in [Fig. 2](#) in order to highlight the different roles that predictability and predictiveness have. The horizontal axis presents the predictability of each TOS category, while the vertical axis presents the predictiveness for each TOS category. The boundary for services to be over and underprovided is shown as a rectangular hyperbola corresponding to the value of the selectivity index for the median service from [Table 2](#). Many of the TOS categories that are the most highly predictable have relatively low predictiveness, with hospice care spending and inpatient visits being the most predictable. At the other extreme, spending on anesthesia is relatively predictive (since anesthesia is almost always done on people who are hospitalized), but has the lowest predictability (it is difficult to foresee the need for anesthesia). Taken together, spending on anesthesia is not vulnerable to service underprovision. Spending on eye procedures has very low predictability and predictiveness, which seems plausible given that expensive eye procedures would rarely be done on persons with other major illnesses, and are usually not repeated in multiple years.

To demonstrate that the results are robust to alternative econometric specifications, [Table 3](#) presents the selection indices for each of the five different econometric specifications, all using TOS as the information set and for disaggregating total spending (full results are available in appendix [Tables A1–A3](#)). The magnitude of the selection indices varies only modestly across specifications, and the ranking of which services are most vulnerable is highly robust to the choice of model. Correlation in these indices across estimation methods is 0.985–0.988. We repeated this sensitivity analysis for decomposed spending by provider specialty (see appendix [Tables A4–A6](#)), with similar findings, and focus in the text on the weighted least squares results for the provider specialty decomposition.

5.4. Results by provider specialty (PS)

[Table 4](#) and [Fig. 3](#) repeat the analysis using spending by provider specialty with the information set being age and gender dummies together with lagged spending by provider specialty. Spending

Table 3
Selection indices by type of service categories for four alternative specifications

Spending by type of service categories	Selection indices				Rankings			
	(1) Original WLS	(2) OLS on full year eligibles	(3) Square root (heter)	(4) Two part linear model	(5) Original WLS	(6) OLS on full year eligibles	(7) Square root (heter)	(8) Two part linear model
Hospice	2.578	4.419	4.304	4.596	1	1	1	1
Home health care	0.875	0.890	0.878	0.950	2	2	2	2
Durable medical equipment	0.703	0.802	0.812	0.842	3	4	3	4
Inpatient visits	0.592	0.816	0.798	0.932	4	3	4	3
Other (including unclassified)	0.495	0.487	0.510	0.500	5	6	5	6
Intermediate care	0.408	0.628	0.486	0.711	6	5	6	5
Hospital visit	0.356	0.392	0.399	0.409	7	8	8	8
Home visit	0.348	0.405	0.454	0.430	8	7	7	7
Inpatient facility (R&B)	0.267	0.304	0.311	0.312	9	10	10	10
Emergency room	0.265	0.327	0.336	0.345	10	9	9	9
Consultation	0.219	0.251	0.253	0.259	11	12	12	12
Other facility services	0.172	0.256	0.281	0.287	12	11	11	11
Outpatient visit	0.170	0.191	0.207	0.199	13	14	17	13
Advanced imaging CAT	0.169	0.192	0.231	0.198	14	13	15	14
Oncology	0.159	0.166	0.163	0.166	15	18	23	18
Lab tests	0.144	0.184	0.245	0.187	16	16	13	17
Other tests	0.134	0.187	0.238	0.194	17	15	14	15
Standard imaging	0.119	0.150	0.175	0.152	18	20	20	21
Specialist	0.114	0.180	0.228	0.192	19	17	16	16
Echography	0.113	0.153	0.177	0.160	20	19	18	19
Ambulatory procedures	0.105	0.134	0.134	0.144	21	24	27	23
Imaging procedure	0.102	0.145	0.154	0.158	22	21	24	20
Office visit	0.096	0.142	0.170	0.144	23	22	21	22
Major proc cardiovascular	0.096	0.128	0.165	0.137	24	25	22	26
Minor procedures	0.095	0.137	0.175	0.142	25	23	19	25
Anesthesia	0.092	0.120	0.125	0.123	26	28	28	28
Endoscopy	0.087	0.113	0.121	0.119	27	29	29	29
Major procedure	0.087	0.102	0.119	0.104	28	30	30	30
Major proc orthopedic	0.086	0.127	0.139	0.133	29	26	25	27
Advanced imaging – MRI	0.083	0.122	0.138	0.143	30	27	26	24
Eye procedures	0.045	0.072	0.069	0.073	31	31	31	31
Correlation with original indexes		0.986	0.985	0.988				

See text for details on estimation and sample selection. Type of service categories are aggregates of the Berenson–Eggers values.

Table 4
Predictability and predictiveness of spending by provider specialty weighted least squares results including partial year eligibles (*N* = 1,380,863)

Spending by provider specialty	Mean (1)	Predictability		Predictiveness (4) Corr(\hat{m}_s , <i>M</i>)	Selection index (5) (3) × (4)
		(2) <i>R</i> ²	(3) CV(\hat{m}_s)		
Pulmonary diseases	25.52	0.080	2.690	0.190	0.511
Oncology	61.82	0.191	6.290	0.074	0.464
Ambulance services	52.77	0.114	2.044	0.220	0.450
Psychiatry	26.05	0.391	5.208	0.058	0.304
Medical specialists	45.82	0.098	2.085	0.144	0.299
Emergency medicine	20.63	0.119	1.390	0.203	0.282
Multispecialty clinic or group	42.88	0.168	3.513	0.077	0.270
Primary care services	230.33	0.187	1.091	0.207	0.226
Podiatry	24.36	0.237	1.860	0.102	0.190
Median service by provider specialty					
Laboratory	55.33	0.458	2.024	0.093	0.188
Urology	52.59	0.425	4.578	0.039	0.181
Radiology	79.64	0.057	0.820	0.206	0.169
Cardiology	96.82	0.090	1.202	0.126	0.151
Neurology and Neurosurgery	26.62	0.029	1.351	0.106	0.144
General surgery	49.92	0.019	0.754	0.187	0.141
Rheumatology	6.54	0.365	5.952	0.023	0.135
Surgical specialists	38.01	0.008	0.729	0.131	0.095
Otolaryngology	12.83	0.052	1.725	0.054	0.093
Orthopedic surgery	54.44	0.022	0.781	0.088	0.068
Ophtho and optometry	96.05	0.050	0.721	0.077	0.055
Dermatology	25.34	0.127	2.268	0.024	0.054
Gynecology	9.41	0.104	4.786	0.005	0.024
Chiropractic	7.66	0.229	3.231	−0.010	−0.032

All models use disaggregated 1996 spending by Part B provider specialty and Part A types of service to predict 1997 spending. Services with less than \$5 of spending omitted. Weighted least squares regression includes all individuals eligible for Medicare Part A and Part B for all of 1996 and for any part of 1997, weighting for fraction of the year eligible in 1997.

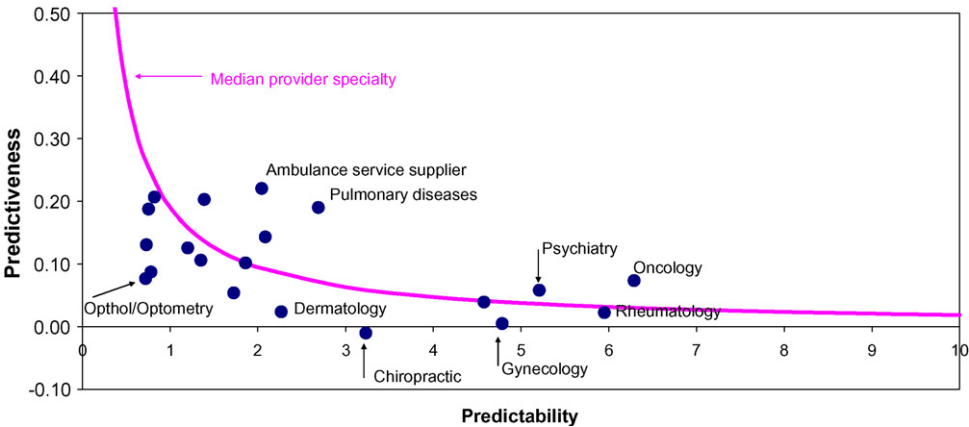


Fig. 3. Plot of predictability vs. predictiveness of spending by provider specialties.

by provider specialty is the appropriate partition of total spending to focus on if the primary mechanism for health plans to influence enrollee selection is by the choice of providers to contract with. The most highly predictable spending by provider specialty using the CV are oncology, psychiatry, and urology. These provider types do not have as much variability relative to their mean as some other categories, however, and hence have more modest predictability indices. The four provider specialties with the highest selection index are pulmonary disease, oncology, ambulance services, and psychiatry. At the other extreme are chiropractic care, gynecology, dermatology, and ophthalmology/optometry, which are services plausibly associated with higher rates of use among healthy rather than unhealthy Medicare beneficiaries. Chiropractic care is interesting in that it is the one provider specialty that is negatively correlated with contemporaneous total spending ($\rho = -0.010, p < 0.001$). Our conceptual model suggests that it would be an attractive service for an HMO to oversupply in that people who use this service have (slightly) lower average spending on total health care.¹⁸

Figs. 2 and 3 reveal some interesting patterns. First, services that are predictable are not always predictive, and both factors matter for the total selection incentive. In most cases, highly predictable services are vulnerable to selection based service underprovision, but not always. Second, spending by TOS in general has higher selection indices, suggesting that selection using type of service will be more effective than selection using provider specialty when identifying services to underprovide. On the opposite side, spending on certain provider specialties is more effective at attracting relatively healthy, low cost enrollees relative to spending on certain types of services. One last contrast is between oncology as a type of service and spending on oncology as a provider specialty. Whereas spending on oncology as a type of service is often done on healthy people for screening and hence is not associated with an incentive to undersupply, spending on oncologists is mostly done on those with confirmed malignancies, and hence is subject to a strong underprovision incentive.

The results in these tables and figures are informative about selection incentives, but not completely informative. Demand elasticities also matter. Our data do not enable us to estimate elasticities at the disaggregated service levels that we consider here, and we are not aware of any empirical estimates by others that have done so. Previous empirical work has shown that certain services, such as inpatient services, have more inelastic demand than total spending, while others, such as psychiatry services, tend to have less inelastic demand. In our framework, this would increase the attractiveness of selection based on psychiatry services, and reduce the attractiveness of selection based on inpatient services. In terms of our figures, higher elasticities work to stretch each point away from the solid line we characterize as implying neutral selection incentives.

6. Discussion

This paper argues that the concept of “predictability” of health care spending needs refinement, conceptually and empirically, in light of health plans’ ability and incentives to ration at the service level to influence selection. Based on what a plan would do to maximize profit, we develop a simple empirical index that can be used to identify services that combine being predictable with

¹⁸ We should highlight that our framework does not take into account any possible demand side moral hazard that might come from increasing the quantity of this service offered. It is possible that if more chiropractic services are offered, then this could reverse the negative correlation we observe given its existing FFS coverage, which is quite limited.

being predictive—the pair of attributes that makes a service vulnerable to underprovision in a capitation-based managed care plan. On one extreme, we find types of services such as hospice care, home health care, and durable medical equipment, and provider specialties of pulmonary care, oncology ambulance, and psychiatry to be subject to underprovision. At the other end of the spectrum, we find services such as eye procedures and MRIs and provider specialties such as chiropractic, gynecology and dermatology to be candidates for overprovision. The ranking of services by this index is very robust to the choice of econometric specification for a given information set.

These results are broadly consistent with the findings of Frank et al. (2000) who analyzed spending patterns by major diagnostic categories in a Medicaid sample and found mental health services to be the most vulnerable to underprovision. Our selection scores are also consistent with the services identified in Cao and McGuire (2003) as being tightly or loosely rationed by Medicare HMOs. Among 11 Part B services that they find to be statistically significantly different from the average in terms of HMO incentive to over- or underprovide, they also find that ambulance service suppliers, psychiatry, and geriatric medicine to be subject to underprovision and chiropractic, ophthalmology/optometry and orthopedic surgery to be subject to overprovision. The one noticeable difference is that they find rheumatology to be subject to underprovision, while we find it to be (weakly) subject to overprovision. Together with these earlier analyses, we have added to the growing literature that identifies sets of services that are most vulnerable to distortion in order to influence service selection.

In our analysis plans choose rationing in response to profit incentives. Payers or regulators may be able to specify minimum levels of intensity for some services (e.g., two day hospital stays for deliveries) overriding a plan's incentive. Our index would not apply to regulated services, but the index would still work to characterize incentives to ration for services under the plan's control.

Identification of the services likely to be subject to too tight or too loose rationing in managed care plans can complement efforts to improve quality of care in managed care plans. Identification of potential problem areas is one form of this complementarity. Another is to design payment systems to plans that attempt to neutralize incentives to ration some services more tightly than others. The incentive to attract or deter potential enrollees based on profitability creates an inefficiency that risk adjustment can be designed to correct. Pairing the insights from the literature on “optimal risk adjustment” with the empirical findings from this and other papers should lead to improvements in risk adjustment policy to counter incentives to over and underprovide services.

Acknowledgements

Research support from grants R01 MH59254 from the National Institute of Mental Health and P01 HS10803 from the Agency for Healthcare Research and Quality is gratefully acknowledged. We are grateful to David Bradford, Jim Burgess, Richard Frank, Jacob Glazer, Will Manning, Joseph Newhouse, and two anonymous referees for helpful comments on earlier versions of this paper and to Zhun Cao for her work as a research assistant. All errors, opinions, and omissions are our responsibility.

Appendix A

See Tables A1–A6 .

Table A1

Predictability and predictiveness of spending by type of service OLS results using only 12 month eligibles ($N = 1,273,471$)

Spending by type of service categories	Mean	Predictability		Predictiveness	Selection index
	(1)	(2) R^2	(3) $CV(\hat{m}_s)$	(4) $\text{Corr}(\hat{m}_s, M)$	(5) $(3) \times (4)$
Hospice	4.55	0.631	58.785	0.075	4.419
Home health care	507.62	0.274	3.050	0.292	0.890
Durable medical equipment	187.02	0.362	3.509	0.229	0.802
Inpatient visits	44.88	0.293	7.089	0.115	0.816
Other (including unclassified)	101.43	0.131	3.453	0.141	0.487
Intermediate care	15.47	0.014	4.654	0.135	0.628
Hospital visit	102.68	0.073	1.103	0.355	0.392
Home visit	22.97	0.322	2.867	0.141	0.405
Inpatient facility (R&B)	2576.00	0.064	0.820	0.371	0.304
<i>Median service by type of service</i>					
Emergency room	24.35	0.168	1.210	0.270	0.327
Consultation	50.96	0.079	0.764	0.328	0.251
Other facility services	84.91	0.070	1.604	0.160	0.256
Outpatient visit	505.08	0.056	0.782	0.244	0.191
Advanced imaging CAT	14.68	0.069	1.232	0.156	0.192
Oncology	19.41	0.015	1.662	0.100	0.166
Lab tests	71.04	0.273	0.966	0.191	0.184
Other tests	33.98	0.138	1.047	0.179	0.187
Standard imaging	45.12	0.091	0.689	0.217	0.150
Specialist	59.42	0.227	1.600	0.113	0.180
Echography	33.68	0.102	0.899	0.171	0.153
Ambulatory procedures	36.10	0.035	0.839	0.160	0.134
Imaging procedure	7.82	0.019	0.901	0.161	0.145
Office visit	159.52	0.333	0.745	0.190	0.142
Major proc cardiovascular	52.34	0.016	0.873	0.146	0.128
Minor procedures	60.07	0.136	1.006	0.136	0.137
Anesthesia	33.76	0.027	0.524	0.229	0.120
Endoscopy	45.27	0.042	0.743	0.152	0.113
Major procedure	35.98	0.008	0.568	0.179	0.102
Major proc orthopedic	32.32	0.010	0.703	0.181	0.127
Advanced imaging – MRI	14.48	0.027	1.117	0.110	0.122
Eye procedures	79.32	0.019	0.622	0.115	0.072
			Correlation with original		0.986

Sample contains individuals eligible for Medicare Part A and Part B for all of 1996 and 1997. Type of service categories are aggregates of the Berenson–Eggers values.

Table A2

Predictability and predictiveness of spending by type of service results using square root model, heteroskedasticity-corrected ($N = 1,273,471$)

Spending by type of service categories	Mean	Predictability		Predictiveness	Selection index
	(1)	(2) R^2	(3) $CV(\hat{m}_s)$	(4) $\text{Corr}(\hat{m}_s, M)$	(5) $(3) \times (4)$
Hospice	4.55	0.627	59.268	0.073	4.304
Home health care	507.62	0.289	2.957	0.297	0.878
Durable medical equipment	187.02	0.356	3.882	0.209	0.812
Inpatient visits	44.88	0.285	7.924	0.101	0.798

Table A2 (Continued)

Spending by type of service categories	Mean (1)	Predictability		Predictiveness (4) Corr(\hat{m}_s, M)	Selection index (5) (3) × (4)
		(2) R^2	(3) CV(\hat{m}_s)		
Other (including unclassified)	101.43	0.138	3.466	0.147	0.510
Intermediate care	15.47	0.013	4.516	0.108	0.486
Hospital visit	102.68	0.073	1.127	0.354	0.399
Home visit	22.97	0.216	5.117	0.089	0.454
Inpatient facility (R&B)	2576.00	0.062	0.851	0.366	0.311
<i>Median service by type of service</i>					
Emergency room	24.35	0.177	1.430	0.235	0.336
Consultation	50.96	0.076	0.816	0.310	0.253
Other facility services	84.91	0.071	1.835	0.153	0.281
Outpatient visit	505.08	0.057	0.812	0.255	0.207
Advanced imaging CAT	14.68	0.075	1.215	0.190	0.231
Oncology	19.41	0.015	1.615	0.101	0.163
Lab tests	71.04	0.189	1.938	0.126	0.245
Other tests	33.98	0.108	1.833	0.130	0.238
Standard imaging	45.12	0.080	0.781	0.224	0.175
Specialist	59.42	0.214	2.225	0.103	0.228
Echography	33.68	0.087	1.112	0.159	0.177
Ambulatory procedures	36.10	0.036	0.864	0.155	0.134
Imaging procedure	7.82	0.019	0.912	0.168	0.154
Office visit	159.52	0.286	1.021	0.166	0.170
Major proc cardiovascular	52.34	0.016	0.884	0.187	0.165
Minor procedures	60.07	0.113	1.513	0.116	0.175
Anesthesia	33.76	0.026	0.534	0.235	0.125
Endoscopy	45.27	0.041	0.781	0.155	0.121
Major procedure	35.98	0.008	0.566	0.211	0.119
Major proc orthopedic	32.32	0.009	0.704	0.198	0.139
Advanced imaging – MRI	14.48	0.028	1.095	0.126	0.138
Eye procedures	79.32	0.018	0.641	0.107	0.069
			Correlation with original		0.985

Sample contains individuals eligible for Medicare Part A and Part B for all of 1996 and 1997. Type of service categories are aggregates of the Berenson–Eggers values. Squared fitted residuals regressed on predicted square root and added to squared prediction for heteroskedasticity correction. Services with less than \$5 of spending omitted.

Table A3

Predictability and predictiveness of spending by type of service results using two part linear model ($N = 1,273,471$)

Spending by type of service categories	Mean (1)	Predictability		Predictiveness (4) Corr(\hat{m}_s, M)	Selection index (5) (3) × (4)
		(2) R^2	(3) CV(\hat{m}_s)		
Hospice	4.55	0.652	61.514	0.075	4.596
Home health care	507.62	0.289	3.199	0.297	0.950
Durable medical equipment	187.02	0.360	3.771	0.223	0.842
Inpatient visits	44.88	0.276	8.400	0.111	0.932
Other (including unclassified)	101.43	0.135	3.591	0.139	0.500
Intermediate care	15.47	0.023	7.691	0.092	0.711
Hospital visit	102.68	0.071	1.180	0.346	0.409
Home visit	22.97	0.321	2.998	0.144	0.430
Inpatient facility (R&B)	2576.00	0.061	0.861	0.363	0.312

Table A3 (Continued)

Spending by type of service categories	Mean	Predictability		Predictiveness	Selection index
	(1)	(2) R^2	(3) $CV(\hat{m}_s)$	(4) $Corr(\hat{m}_s, M)$	(5) $(3) \times (4)$
<i>Median service by type of service</i>					
Emergency room	24.35	0.175	1.276	0.270	0.345
Consultation	50.96	0.077	0.809	0.321	0.259
Other facility services	84.91	0.071	2.182	0.132	0.287
Outpatient visit	505.08	0.057	0.830	0.240	0.199
Advanced imaging CAT	14.68	0.077	1.436	0.138	0.198
Oncology	19.41	0.015	1.973	0.084	0.166
Lab tests	71.04	0.270	0.985	0.190	0.187
Other tests	33.98	0.136	1.093	0.177	0.194
Standard imaging	45.12	0.090	0.719	0.212	0.152
Specialist	59.42	0.229	1.695	0.113	0.192
Echography	33.68	0.102	0.953	0.168	0.160
Ambulatory procedures	36.10	0.035	0.991	0.145	0.144
Imaging procedure	7.82	0.016	1.087	0.145	0.158
Office visit	159.52	0.325	0.752	0.191	0.144
Major proc cardiovascular	52.34	0.014	0.993	0.138	0.137
Minor procedures	60.07	0.133	1.070	0.133	0.142
Anesthesia	33.76	0.026	0.554	0.223	0.123
Endoscopy	45.27	0.040	0.812	0.147	0.119
Major procedure	35.98	0.008	0.610	0.171	0.104
Major proc orthopedic	32.32	0.008	0.811	0.164	0.133
Advanced imaging – MRI	14.48	0.026	1.399	0.102	0.143
Eye procedures	79.32	0.018	0.638	0.115	0.073
			Correlation with original		0.986

Sample contains individuals eligible for Medicare Part A and Part B for all of 1996 and 1997. Type of service categories are aggregates of the Berenson–Eggers values.

Table A4

Predictability and predictiveness of spending by provider specialty OLS results using only 12 month eligibles ($N = 1,273,471$)

Spending by provider specialty	Mean (1)	Predictability		Predictiveness (4) Corr(\hat{m}_s, M)	Selection index (5) (3) × (4)
		(2) R^2	(3) CV(\hat{m}_s)		
Pulmonary diseases	18.57	0.099	2.572	0.168	0.432
Oncology	48.54	0.103	4.858	0.071	0.347
Ambulance services	39.90	0.110	2.056	0.258	0.529
Psychiatry	25.17	0.301	4.662	0.101	0.471
Medical specialists	39.08	0.070	1.649	0.168	0.276
Emergency medicine	17.51	0.139	1.435	0.228	0.326
Multispecialty clinic or group	38.49	0.093	2.536	0.092	0.232
Primary care services	206.57	0.159	0.919	0.255	0.234
Podiatry	23.70	0.241	1.901	0.131	0.250
<i>Median service by provider specialty</i>					
Laboratory	52.87	0.166	1.249	0.174	0.217
Urology	51.26	0.430	4.618	0.046	0.211
Radiology	71.71	0.060	0.783	0.229	0.179
Cardiology	89.99	0.093	1.177	0.145	0.171
Neurology and Neurosurgery	24.28	0.024	1.196	0.151	0.181

Table A4 (Continued)

Spending by provider specialty	Mean	Predictability		Predictiveness	Selection index
		(2)	(3)		
	(1)	R^2	$CV(\hat{m}_s)$	(4) $Corr(\hat{m}_s, M)$	(5) $(3) \times (4)$
General surgery	44.75	0.022	0.737	0.200	0.147
Rheumatology	6.50	0.337	5.514	0.039	0.213
Surgical specialists	35.23	0.009	0.733	0.145	0.106
Otolaryngology	12.63	0.053	1.686	0.068	0.114
Orthopedic surgery	53.23	0.025	0.817	0.130	0.106
Ophtho and optometry	96.78	0.051	0.723	0.108	0.078
Dermatology	25.67	0.125	2.230	0.034	0.075
Gynecology	9.42	0.062	3.389	0.009	0.032
Chiropractic	7.86	0.232	3.207	−0.003	−0.010
			Correlation with original		0.924

All models use disaggregated 1996 spending by Part B provider specialty and Part A types of service to predict 1997 spending. Services with less than \$5 of spending omitted 12 month eligible sample contains only individuals eligible for Medicare Part A and Part B for all of 1996 and 1997.

Table A5

Predictability and predictiveness of spending by provider specialty results using square root model, heteroskedasticity-corrected ($N=1,273,471$)

Spending by provider specialty	Mean	Predictability		Predictiveness	Selection index
		(2)	(3)		
	(1)	R^2	$CV(\hat{m}_s)$	(4) $Corr(\hat{m}_s, M)$	(5) $(3) \times (4)$
Pulmonary diseases	18.57	0.084	3.870	0.136	0.528
Oncology	48.54	0.104	5.264	0.073	0.383
Ambulance services	39.90	0.121	2.227	0.244	0.543
Psychiatry	25.17	0.276	6.449	0.089	0.573
Medical specialists	39.08	0.077	2.096	0.149	0.311
Emergency medicine	17.51	0.108	2.377	0.163	0.387
Multispecialty clinic or group	38.49	0.072	4.349	0.072	0.314
Primary care services	206.57	0.151	1.250	0.226	0.283
Podiatry	23.70	0.152	4.530	0.083	0.374
<i>Median service by provider specialty</i>					
Laboratory	52.87	0.135	2.742	0.116	0.319
Urology	51.26	0.451	4.917	0.047	0.234
Radiology	71.71	0.051	0.928	0.238	0.221
Cardiology	89.99	0.074	1.628	0.131	0.213
Neurology and Neurosurgery	24.28	0.022	1.454	0.156	0.227
General surgery	44.75	0.021	0.812	0.188	0.153
Rheumatology	6.50	0.243	12.385	0.026	0.323
Surgical specialists	35.23	0.008	0.757	0.164	0.124
Otolaryngology	12.63	0.046	2.411	0.060	0.144
Orthopedic surgery	53.23	0.022	0.933	0.116	0.108
Ophtho and optometry	96.78	0.047	0.834	0.102	0.085
Dermatology	25.67	0.104	3.690	0.024	0.088
Gynecology	9.42	0.065	2.982	0.009	0.027
Chiropractic	7.86	0.139	6.159	0.000	0.002
			Correlation with original		0.889

All models use disaggregated 1996 spending by Part B provider specialty and Part A types of service to predict 1997 spending. Dependent variable is square root of spending. Squared fitted residuals regressed on predicted square root and added to squared prediction for heteroskedasticity correction. Services with less than \$5 of spending omitted 12 month eligible sample contains only individuals eligible for Medicare Part A and Part B for all of 1996 and 1997.

Table A6

Predictability and predictiveness of spending by provider specialty results using two part linear model ($N = 1,273,471$)

Spending by provider specialty	Mean (1)	Predictability		Predictiveness (4) $\text{Corr}(\hat{m}_s, M)$	Selection index (5) $(3) \times (4)$
		(2) R^2	(3) $\text{CV}(\hat{m}_s)$		
Pulmonary diseases	18.57	0.089	2.977	0.164	0.488
Oncology	48.54	0.101	5.501	0.068	0.372
Ambulance services	39.90	0.112	2.332	0.241	0.562
Psychiatry	25.17	0.291	5.152	0.102	0.525
Medical specialists	39.08	0.067	1.853	0.169	0.313
Emergency medicine	17.51	0.138	1.574	0.221	0.347
Multispecialty clinic or group	38.49	0.088	2.784	0.089	0.248
Primary care services	206.57	0.157	0.948	0.255	0.242
Podiatry	23.70	0.232	2.064	0.125	0.259
<i>Median service by provider specialty</i>					
Laboratory	52.87	0.163	1.320	0.173	0.228
Urology	51.26	0.441	4.947	0.048	0.238
Radiology	71.71	0.058	0.832	0.223	0.186
Cardiology	89.99	0.092	1.220	0.148	0.181
Neurology and Neurosurgery	24.28	0.020	1.380	0.140	0.193
General surgery	44.75	0.020	0.806	0.193	0.155
Rheumatology	6.50	0.323	6.104	0.040	0.245
Surgical specialists	35.23	0.006	0.818	0.132	0.108
Otolaryngology	12.63	0.045	1.969	0.064	0.126
Orthopedic surgery	53.23	0.022	0.854	0.129	0.110
Ophtho and optometry	96.78	0.049	0.756	0.105	0.080
Dermatology	25.67	0.119	2.496	0.035	0.088
Gynecology	9.42	0.062	3.874	0.009	0.037
Chiropractic	7.86	0.218	3.447	−0.001	−0.003
			Correlation with original		0.922

All models use disaggregated 1996 spending by Part B provider specialty and Part A types of service to predict 1997 spending. First stage uses Probit, second stage uses OLS. Smearing used to predict true mean correctly. Services with less than \$5 of spending omitted 12 month eligible sample contains only individuals eligible for Medicare Part A and Part B for all of 1996 and 1997.

References

- Buntin, M.L., Zaslovsky, A.M., 2004. Too much ado about two part models and retransformation? Comparing methods of modeling Medicare expenditures. *Journal of Health Economics* 23, 525–542.
- Cao, Z., McGuire, T.G., 2003. Service-level selection by HMOs in Medicare. *Journal of Health Economics* 22, 915–931.
- Cao, Z., 2002. HMO selection at the service level: evidence from Medicare plan choice behavior, Chapter 3 of Service-level Risk Selection by HMOs in Medicare, Ph.D. Dissertation, Boston University.
- Cutler, D., 2003. *Your Money or Your Life*. Oxford University Press.
- Ellis, R.P., Ash, A.S., 1995. Refinements to the diagnostic cost group (DCG) model. *Inquiry* 32 (4), 418–429.
- Ellis, R.P., Pope, G.C., Iezzoni, L.I., Ayanian, J.Z., Bates, D.W., Burstin, H., Ash, A.S., 1996. Diagnosis-based risk adjustment for Medicare capitation payments. *Health Care Financing Review* 12 (3), 101–128.
- Feldman, R., Dowd, B., 1982. Simulation of a health insurance market with adverse selection. *Operations Research* 30 (6), 1027–1042.
- Frank, R., Glazer, J., McGuire, T.G., 2000. Measuring adverse selection in managed health care. *Journal of Health Economics* 19, 829–854.
- Glazer, J., McGuire, T.G., 2000. Optimal risk adjustment of health insurance premiums: an application to managed care. *American Economic Review* 90 (4), 1055–1071.

- Glazer, J., McGuire, T.G., 2002. Setting health plan premiums to ensure efficient quality in health care: minimum variance optimal risk adjustment. *Journal of Public Economics* 84, 153–173.
- Keeler, E.B., Newhouse, J.P., Carter, G., 1998. A model of the impact of reimbursement schemes on health plan choice. *Journal of Health Economics* 17 (3), 297–320.
- Luft, H.S., Miller, R.H., 1988. Patient selection and competitive health system. *Health Affairs* 7 (3), 97–112.
- Manning, W.G., 1998. The logged dependent variable, heteroskedasticity, and the retransformation problem. *Journal of Health Economics* 17 (3), 283–296.
- Mello, M.M., Stearns, S.C., Norton, E.C., 2002. Do Medicare HMOs still reduce health services use after controlling for selection bias? *Health Economics* 11, 323–340.
- Miller, R.H., Luft, H.S., 2002. HMO plan performance update: an analysis of the literature, 1997–2001. *Health Affairs* 21 (4), 63–86.
- McCall, N., Wai, H.S., 1983. An analysis of the use of Medicare services by the continuously enrolled aged. *Medical Care* 21 (6), 567–585.
- Newhouse, J.P., Manning, W.G., Keeler, E.M., Sloss, E.M., 1989. Adjusting capitation rates using objective health measures and prior utilization. *Health Care Financing Review* 10 (3), 41–54.
- Newhouse, J.P., 2002. *Pricing the Priceless: A Health Care Conundrum*. MIT Press, Cambridge.
- Newhouse, J.P., 1997. Risk adjustment and Medicare: taking a closer look. *Health Affairs* 13 (1), 132–146.
- Pope, G.C., Ellis, R.P., Ash, A.S., Ayanian, J.Z., Bates, D.W., Burstin, H., Iezzoni, L.I., Marcantonio, E. Wu, B., 2000. Diagnostic cost group hierarchical condition category models for Medicare risk adjustment. Final Report to Health Care Financing Administration, December.
- Pope, G.C., Kautter, J., Ellis, R.P., Ash, A.S., Ayanian, J.Z., Iezzoni, L.I., Ingber, M.J., Levy, J.M., Robst, J., 2004. Risk adjustment of Medicare capitation payments using the CMS-HCC model. *Health Care Financing Review*, Summer 25 (4).
- Shen, Y., Ellis, R.P., 2002a. How profitable is risk selection? A comparison of four risk adjustment models. *Health Economics* 11 (2), 165–174.
- Shen, Y., Ellis, R.P., 2002b. Cost minimizing risk adjustment. *Journal of Health Economics* 21 (3), 515–531.
- van Vliet, R., 1992. Predictability of individual health care expenditures. *Journal of Risk and Insurance* 56 (3), 443–460.
- Veazie, P.J., Manning, W.G., Kane, R.L., 2003. Improving risk adjustment for Medicare capitated reimbursement using nonlinear models. *Medical Care* 41 (6), 741–752.
- Welch, W.P., 1985. Medicare capitation payments to HMOs in light of regression toward the mean in health care costs. In: Scheffler, R.M., Rossiter, L.F. (Eds.), *Advances in Health Economics and Health Services Research*. JAI Press, London.