Does Service-Level Spending Show Evidence of Selection Across Health Plan Types?

Randall P. Ellis[1], Shenyi Jiang[2], Tzu-Chun Kuo[3]

June 17, 2011

[1]Department of Economics, Boston University,270 Bay State Road, Boston, MA 02215
   (email: ellisrp@bu.edu)
[2]Hanqing Advanced Institute of Economics and Finance, School of Finance, Renmin
   University of China, 59 Zhongguancun St., Haidian District, Beijing, China 100872 (email:
   syijiang@gmail.com)
[3]Verisk Health, Inc.,130 Turner Street, Waltham, MA 02110 (email:tkuo@veriskhealth.com)

Does Service-Level Spending Show Evidence of Selection Across Health Plan Types?

We provide an explanation for the widespread finding that capitated managed care plans attract comparatively healthy, low cost enrollees relative to traditional unmanaged plans. Using disaggregated commercial insurance claims from the Thomson-Reuters MarketScan database, we show that managed care plans spend proportionally less on those types of services that are predicted to be more profitable to ration tightly using a selection index developed by Ellis and McGuire that captures the derivative of profits with respect to reduced spending on disaggregated services. Conventional diagnosis-based risk adjusted premiums reduce selection incentives by about 50% relative to premiums that are not risk-adjusted.
(*JEL:* I11, C21, D12)

Fixed premiums create strong profit incentives for health plans to prefer enrolling healthy low-cost rather than sicker, high-cost enrollees, since premiums rarely reflect the full cost differential between sick and healthy enrollees. While governments and employers can carefully choose their benefits, and prohibit selective marketing and explicit exclusion of high-cost enrollees, a recent concern is that managed care plans may nonetheless over- or under- supply certain medical services so as to influence individual enrollment decisions in ways that are difficult to regulate. A series of recent papers has shown that service-level distortions --- i.e. under-provision of services used primarily by the sick and overprovision of services that are used primarily by the healthy --- are potentially profitable. The primary contribution of this paper is that we test the predictions of this literature by examining whether the services identified as profitable to over- (or under-) supply are actually the ones over (or under) supplied in practice. We also examine whether risk adjustment reduces these service distortion incentives. The results of our analysis are important as policymakers and researchers continue to ask whether competition can be relied upon to create appropriate incentives for quality and cost among competing health plans.

Although we frame the problem in this paper entirely as the problem facing competing health plans, the model that we develop is actually quite generic, and applies also

to other marketsin which a bundle of different services are sold at a fixed price, and consumers vary significantly in their costs. Consider, for instance, the following situations: a software company trying to choose between excellent packaging and excellent phone support for their software; an extended appliance repair contract choosing between convenient renewals or convenient repairs; and an automobile insurance company choosing between convenient premium payment options or speedy settlement of accident claims. In each market the first-listed service in each pair is valued by all types of customers, whether low or high cost, while the second dimension of service is valued in particular by high cost consumers. The essence of our model is that in each setting profit maximizing firms will choose to oversupply the first dimension of services and undersupply the second relative to the welfare-optimizing levels.

Recent theoretical and empirical studies in the health care literature have focused on identifying and correcting service-level selection incentives, by which we mean the incentives to influence enrollee types by over- or under-supplying certain health care services. Service distortions are particularly of concern with managed care health plans, since their closer involvement in selecting providers with whom to contract and specifying the constraints under which providers work give them greater ability to influence the services that are over or undersupplied relative to non-managed care plans. In the US, a whole array of health plan types have emerged that differ in the extent to which they manage the services provided to consumers, and this provides a natural setting for examining how firms with alternative management contracts differ in the services they offer. Among the common types of health plans in the US, traditional comprehensive plans (COMP) place the least restrictions on patient choice of providers or choice of services: patients can for the most part visit any provider at any time and will have coverage for almost any services that are covered. At the other extreme are Health Maintenance Organizations (HMOs) which selectively contract with

a subset of doctors and hospitals in an area, and often require ex ante preauthorization or ex post justification of services received. In between these two extremes preferred provider organizations (PPOs) generally use selective contracting with certain but not all providers and generally arrange provider discounts to control costs. Point of service (POS) plans generally combine management services of HMOs with relatively unrestricted access to providers outside of the negotiated provider network, and hence represent a form of managed care that is looser than HMOs but tighter than PPOs or COMP.

In addition to managed care, researchers, policymakers, employers and governments in the US and elsewhere often modify the fixed premiums using experience rating or "risk adjustment" so that the fixed payment (premiums) better reflect the expected cost of plan enrollees. It is of considerable interest to understand how well risk adjustment mitigates existing service distortion incentives, and hence we examine how risk adjustment changes our results below.

The literature of closest relevance to this paper examines provider incentives created by fixed premium payments. Glazer and McGuire (2000) were the first to distinguish between conventional risk adjustment, which pays health plans the expected cost of each enrollee based on observable enrollee characteristics, and optimal risk adjustment, in which plan premiums are adjusted so as to offset service level selection incentives. The Glazer and McGuire solution is to overpay health plans for signals that identify enrollees as high risk and underpay plans for signals that identify the enrollee as low risk. Frank, Glazer and McGuire (2000) extend this framework by explicitly modeling service-level spending decisionsin the absence of optimal risk adjustment, when health plans have private information not observed by the payer. They set up their model such that health plans use shadow prices to decide how extensively to ration each health care service, and demonstrate that differences in selection incentives vary as their model predicts in US Medicaid data.

In a precursor to this paper, Ellis and McGuire (2007) (henceforth EM) make further progress by deriving a relatively straightforward selection index that they implement empirically. Their selection index is the product of two measures --- predictability (how well individuals can predict their subsequent use of each service) and predictiveness (i.e., how well spending on each service predicts plan profitability). Both of these concepts are explained further below. Using Medicare data from 1996 and 1997, EM calculate selection indices and find that services such as hospice care, home health care and durable medical equipment are the most tightly controlled (i.e., underprovided), while services such as eye procedures and magnetic resonance imaging tend to be over provided under capitated payment. Because they only use Medicare data from traditional indemnity health plans, EM do not actually show that plans behave as predicted by the model, only establish that they have an incentive to do so. A number of further empirical studies have demonstrated that managed care plans and non-managed care plans have different levels of spending on various services(e.g., Cao and McGuire, 2003; Eggleston and Bir, 2007), but none of these papers actually test whether the services over- or under-supplied by managed care plans are the services identified by the theoretical literature as most profitable to do so.

This paper is the first to actually test the predictions of the Frank, Glazer, and McGuire model using the EM selection index framework. Using a rich data set on the privately insured sample, we overcome two weaknesses in those papers: first, existing studies have used only Medicare or Medicaid samples which do not represent the privately-insured population who make up the largest share of the US insured population; second, while selection incentives are most relevant for capitated managed care plans, existing studies have considered only non-managed care data. Our 2003 and 2004 privately insured Thomson-Reuters (formerly MEDSTAT) MarketScan commercial claims and encounters database contains a diversified group of enrollees who enroll in a wide array of plans. These include both managed care

health plans and non-managed care plans. This study is the first in which the predictions are tested and extended using both managed and non managed care data.

In the remainder of the paper we answer three questions:

1. Are service-level selection indices for managed care plans similar to those of non-managed care plans? In other words, do all plan types have similar incentives to distort certain services?

2. How well does risk adjustment reduce the incentive of plans to risk select through service distortion?

3. Is the actual pattern of over- and under-provision of certain services in each plan type consistent with the predictions of the selection index model?

The rest of the paper is structured as follows. Section I reviews the Ellis-McGuire Selection Index. The data used in this study is summarized in Section II. Section III describes the estimation strategy. Section IV presents the empirical results. Section V includes brief concluding remarks as well as suggestion for future research.

I.   The Ellis-McGuire (EM) Selection Index

Our estimation is based on the Ellis-McGuire (2007) selection index, which is briefly re-derived in an appendix to this paper. The EM model assumes that health plans offer S services and make actuarially fair total profits. Individuals choose their health plans based on their expected covered spending on each service s, $\hat{m}_s$. It is assumed that individuals are rational and respond to health plans' service-level offerings when choosing plans. Health plans anticipate individuals' reactions, and tighten or loosen the availability of services in order to attract favorable (profitable) individuals and avoid unprofitable ones.

The EM selection index I_s, is the elasticity of health plan profits with respect to one additional dollar spent on a given health care service s. After normalizations, EM the selection index can be written as:

$$I_s = \phi \cdot \sigma_\Pi \eta_s \left(\frac{\sigma_{\hat{m}_s}}{\overline{m}_s} \rho_{\hat{m}_s, \Pi} - C \right) \tag{1}$$

where ϕ is the density of a uniform distribution of individual specific valuation regarding a plan's non-service component. σ_Π is the standard deviation of profit function. η_s is the demand elasticity for services. $\frac{\sigma_{\hat{m}_s}}{\overline{m}_s}$, the variation coefficient of the predicted spending on service s. $\rho_{\hat{m}_s, \Pi}$ is the correlation between predicted service spending $\hat{m}_s$ and the profit Π. The term C is a numeric constant to capture terms that do not depend on service s.

Under the assumption that premiums are constant for all individuals, it is straightforward to rewrite the selection formula to incorporate the correlation coefficient with individual total spending M, instead of profit Π. The standard deviation of total spending is equal to that of profits. Since profits and spending are negatively related, this change flips the sign of the index (1). Therefore, the EM index can be rewritten as

$$I_s = -\phi \cdot \sigma_M \cdot \eta_s \left(\frac{\sigma_{\hat{m}_s}}{\overline{m}_s} \rho_{\hat{m}_s, M} + C \right), \tag{2}$$

The first term in this expression is the demand elasticity of service s, which we do not estimate here but discuss below. ϕ, σ_Π, and C are constant across services. They change the magnitudes but not the orders of the indices, thus we do not include them in the further analyses. Of greater interest is the term $\frac{\sigma_{\hat{m}_s}}{\overline{m}_s}$, the variation coefficient of the predicted spending on service s, which EM define as the "predictability" of service s.[1] It compares the volatility of expected spending on service s to its mean. Smaller volatility implies lower predictability. Intuitively, health plans have strong incentive to ration services with high

---

[1] As in EM, we assume that $\overline{m}_s = \overline{\tilde{m}}_s$. It is also one of the statistical properties of the least square estimation. As we use weighted least square specification to predict next year service level spending, this equation holds in our model.

predictability, not only because they are effective in influencing consumer choices, also large volatility poses greater risk to health plans. The third term, $\rho_{\hat{m}_s,M}$, which EM define as the "predictiveness", is the correlation between predicted service spending $\hat{m}_s$ and total spending M. This term implies that the higher this correlation is for a service, the stronger the incentive to reduce its use. Plans dislike providing services that are positively correlated with total spending but like to provide services that are negatively correlated with total health care spending. In the analysis below, we only present the selection index by the product $\dfrac{\sigma_{\hat{m}_s}}{\overline{m}_s}\rho_{\hat{m}_s,M}$ for ease of interpretation.[2]

Assuming a constant premium is unrealistic given that most health plans charge premiums that adjust at least for age and gender. In many instances premiums are also adjusted using information such as health status or prior utilization information, whether formally or implicitly through experience rating. Under risk adjustment, $\rho_{\hat{m}_s,\Pi}$ and σ_Π are no longer equal to $\rho_{\hat{m}_s,M}$ and σ_M respectively. The changes in these terms are explicitly taken into account when we discuss the effects of risk-adjusted premiums later in the paper.

In order to understand the consequences of risk adjustment, in the analysis below we recalculate the selection indices under three scenarios: no risk adjustment, age and gender risk adjustment, and diagnosis-based risk adjustment. Since the index corresponds to the slope of the profit function, these measures are of direct interest for quantifying the change

---

[2]The above expression for the selection index may seem unintuitive to some readers, but can also be motivated in the following more heuristic way. We are interested in characterizing services for which increased spending is associated with increased profits. We therefore expect these services to have the property that the covariance of profits with spending on these services is positive (providing more of this service increases profits). But consumers do not base their enrollments on realized spending but rather on expected spending. Hence consumer's expected spending on a service should positively covary with total profits. We don't have any way of estimating the demand response of health plan enrollments to spending on aggregate services, so instead of measuring the total covariances, we calculate an index that permits us to quantify the relative covariance of expected spending on each service with total profits. In order to allow this relative comparison, instead of the total covariance, we rearrange the terms so that the two components are unit free, comprising a coefficient of variation and a correlation coefficient. We are unable to estimate the third component of the index, the elasticity of demand, which is also unit free.

incentives from risk adjustment. For the last scenario, we aggregate diagnoses into binary explanatory variables using the Verisk Health hierarchical condition category (HCC) classification system (Ash et al., 2000), which is a richer, more predictive model than the one use for the US for Medicare payments to managed care plans. .

The EM selection index is attractive as both the predictability and predictiveness terms can be empirically estimated. It is clear that η_s, the demand elasticity for services, also affects the profitability of providing more of service *s* since it determines how responsive enrollees are to prices. Unfortunately, without significant variation in cost sharing or service prices, we are unable to estimate the demand elasticity. Previous studies find that empirical estimates of price elasticity of demand do not vary much between inpatient and outpatient services, and between preventive and acute cares (see Ringel, et al, 2005; Manning, et al, 1987; Meyerhoefer, et al, 2009). A recent paper by Duarte (2011) estimates price elasticity of demand for a few selected services, and finds greater price effect in services such as home visits ($\eta_D \approx -1.89$) and psychologist ($\eta_D \approx -2.08$), compared to services such as Appendectomy ($\eta_D \approx -0.07$) or Cholecystectomy($\eta_D \approx -0.05$). In our results below, we find that physician services with higher consumer demand elasticities (e.g. outpatient services) also tend to have greater selection indices than physician services with lower demand elasticities (e.g., most inpatient services), hence incorporating service level demand elasticities would tend to reinforce, not weaken our results. Following EM, we calculate EM selection indices assuming elasticities are the same for services; we return to discuss the significance of these elasticities in our interpretation section.

II.   Data

We use the 2003 and 2004 Thomson-Reuters MarketScan commercial claims and encounters database. This large database contains service-level inpatient and outpatient

medical claims, encounter records, prescription drug claims, enrollment and eligibility information from 45 large employers, health plans, governments and public organizations. For this project, we collapse the claims and encounters data by detailed patient information across sites and types of providers. The advantage of our commercial data compared to the Medicare data, is that it includes data from managed care plans, which are not readily available for Medicare enrollees. This data set also enables us to compare incentives and behaviors across four broad health plan types.

We focus our analysis on working adults with single plan coverage (the data does not allow us to identify whether these workers were single or chose individual over family coverage, only that there was one person in the contract), aged 21 to 64, including partial eligible enrollees in 2004. Our final sample includes 3.29 million individuals.

In order to calculate the selection index, we predict 2004 service-level spending based on information from 2003, the previous year. Our information set includes previous year's spending in each service, prior year diagnoses, age and gender information. Ideally we would like to partition spending to be consistent with what health plans actually use to influence their service provisions. However, little is known about how health plans ration services in practice. Conceptually, health plans might have different coverage for different procedures. In the extreme case, some managed care plans restrict the use of certain types of services, which suggests that selection happens for each type of service, and hence this is the most appropriate decomposition of spending. To partition spending by type of service, we classify the procedure codes for each outpatient and inpatient claim using the widely-used Berenson-Eggers type of service codes.We then further aggregate these clusters of services into 25 categories. Since non-phyisican services are not classified in the Berenson-Egger system, for

claims with missing procedure codes, we used place of service codes[3] and group them into 8 categories:, including inpatient facility, office/clinic/ambulatory center, skilled nursing facility/intermediate care, hospice, home/mobile clinic, and pharmacy. In total, we create 33 type of service categories.

We then aggregate spending from detailed outpatient, inpatient and drug claims into their type of service categories for each enrollee. The key spending variable we focus on is the covered charge, a financial variable on claims that best approximates the medical resources used in treating patients. We are aware that this variable does not try to capture the incentives of out-of-pocket payment by patients, nor does it take into account provider bonuses or discounts, but it is a widely used measure without clearly superior alternatives.

III.   Estimation Strategy

To obtain the EM selection index for each service s, we need to separately calculate predictability and predictiveness, which requires predicting service level spending for the prediction year (2004) using information from the prior year (i.e., 2003).

We first consider the choice of an econometric specification. Although various advanced econometric methods, including generalized linear model specifications and two-part models using linear and transformed expenditures have been developed to overcome the classical problems of large proportion of zero expenditures and long right trail of health expenditures, several recent studies (EM, 2007, and Dusheiko et. al., 2009) have shown that with very large samples, ordinary least squares model (OLS) performs about as well as more advanced econometric specifications at recovering predicted subsample means. We test the model performance by using 2003 information to predict 2004 annual spending and pick the one with the highest explanatory power in terms of R^2. In Appendix A we show the results

---

[3] Some services are billed only by providing the service place, regardless of what kind of procedure is conducted, which results in missing procedure codes.

using three alternative econometric specifications: weighted least squares (weighted LS) using annualized covered health spending weighted by the fraction of the year eligible, OLS and two-part linear models. For the OLS and the two-part linear models, partial year enrollees are excluded. The R^2 for these three specifications are very similar, which confirms EM's finding. For the rest of our analysis, we focus on the weighted LS model, which is essentially an OLS with samples including partial eligible enrollees.

We then turn to the information sets used for predictions. We test several combinations, all of which include 10 age-gender dummy variables. Besides age and gender groups, the other explanatory variables include one of the following, prior year total covered charges, diagnosis categories, or prior year service-level spending decomposed by type of service. For the diagnosis-based models we use the hierarchical condition category classification system (Ash et al., 2000). Following EM, we found the most predictive information set to use to predict 2004 spending by type of services was disaggregated spending by type of service the preceding year. Hence to approximate consumer expectations of spending by type of service we ran regressions for each type of service predicting the 2004 spending using discrete age and gender groups and the full array of 33 spending variables by type of service in 2003. In regression terms, we estimated the following regression equation via ordinary least squares for each of the s services.

$$m_s^{2004} = f(agegender, m_1^{2003}, ..., m_{33}^{2003}) .$$

Predictability and predictiveness measures are calculated using the predictions from this regression, namely the coefficient of variation of the expected spending and then the sample correlation between this expected spending measure and the relevant measure of profits for each assumed level of risk adjustment. The selection index is reported as the product of these two terms $\frac{\sigma_{\hat{m}_s}}{\bar{m}_s} \rho_{\hat{m}_s, M}$. We first calculate these indices using a pooled

regression with data from all plan types. We then run new regressions and recalculate the selection indices separately for each plan type.[4] Comparisons of rankings and magnitudes of indices allow us to see whether various plan types have similar incentives to distort certain services, which is the first goal of our analysis.

Our second goal is to see how incentives differ between models with flat premiums and premiums that vary by the extent of risk adjustment. In addition to the no risk adjustment results, we also calculate selection indices after premiums are risk adjusted. If $\Pi = \text{Premium} - M$, then $\rho_{\hat{m}_s, \Pi} = -\rho_{\hat{m}_s, M}$ under the condition that premiums are constant. However, the equality is no longer valid if individuals pay risk adjusted premiums. As we do not directly observe premiums in our data, we need to approximate the premiums. We try three ways of risk adjustment. The first assumes no risk adjustment, the second uses only age and sex information, and the third one includes age, sex and diagnostic information captured by HCC in 2003. With predicted revenue imputed from the each assumption of how premiums are risk adjusted, we calculate $\tilde{\Pi}$, and hence $\rho_{\hat{m}_s, \tilde{\Pi}}$.

Our third goal is to see if the pattern of actual service-level spending is consistent with the predictions by selection indices. To do this, we both test the associations statistically and examine patterns graphically by plan type. For both analyses we create a new variable that we call the MEANRATIO. MEANRATIO for each plan and each service is calculated as the percent of total plan spending on a given service in a given health plan type divided by percent of spending on that same service for all plan types. Hence if HMOs spend 15 percent of their health spending on service XYZ, while the average for all plan types on this type of service is 10 percent, then the mean ratio would be 15/10 = 1.5. A PPO plan spending 5

---

[4] We choose 4 plan types with the highest market share, accounting for over 96 percent enrollment in year 2004 in the Marketscan data. They are PPO (51.6%), POS (19.8%), COMP (17.5%) and HMO (7.3%). We drop basic/major medical (0%), point of service with capitation (3.8%) and exclusive provider organizations (0.02%) because of their small sample sizes.

percent of its total on XYZ would have a mean ratio of 5/10=.5. The reasons for this normalization in our mean ratio are twofold. The first one is that some expensive services have small means due to extremely low frequency. Without normalization, services with large means affect profits more in dollar terms, but perhaps less in proportion to the mean. The second reason is that this normalization removes systematic price differences by plan type. If PPOs negotiate discounts across all services relative to COMP and this is reflected in premiums, then this does not affect selection incentives or our MEANRATIO variable.

IV.   Empirical Results

Table 1 provides summary statistics for four types of plans: COMP, PPO, POS and HMO plans on mean age and total covered health spending[5]. Average spending increased between 2003 and 2004 for all plan types. Comprehensive spending increased the most. HMO enrollees are, on average, younger than other plan enrollees. If we compare HMO with COMP, HMO plan enrollees are on average about 8 years younger and are associated with approximately two thirds of comprehensive plan's average total spending. Average total spending and mean age for PPO and POS enrolleeslies in between COMP and HMO. The large difference in average spending across plans reflects the variations in enrollees' health status, as well as in health plan generosity.[6] Due to adverse selection, older and less healthy consumers prefer more generous health plan options such as COMP and PPO rather than POS and HMO plans.

A. *Predictive Power of Different Information Sets for Total Spending*

---

[5] We calculate annualized spending for year 2004, weighted by fraction of the year enrolled in a plan type.
[6] An earlier draft of this paper used a larger sample of HMO enrollees, however it was found that HMO enrollees in California had implausibly low annual spending estimates, perhaps due to different method of assigning prices to procedures and patients in staff model HMOs that are common in California. We omitted all California HMO enrollees from the current analysis. Our empirical results using the larger dataset (including California) were very similar to those reported here using this restricted sample.

We first consider the predictive power of alternative econometric specifications (Appendix Table A1). Three econometric specifications, weighted LS, OLS, and two-part linear model regressions, all perform well and have similar R^2 when explanatory variables are the same. The two-part linear model does not show clear superiority over simple OLS.[7] As in EM, we focus the remainder of our analysis on weighted least square results.

We next compare the predictive power of different information sets, holding the econometric model fixed (Appendix Table A1). With age/gender as the sole predictor, R^2 is less than 2 percent. Controlling for prior year total spending along with age/gender, R^2 increases to 8-9 percent. Instead of using prior year total spending, we use individual diagnostic information as summarized in the HCC dummies, the R^2 improves to 16-17 percent. Finally, using disaggregated spending by type of service for the prior year results in an R^2 of 20.3%. Similar results are found for spending on each disaggregated service. As in EM, for the remainder of our analysis we focus on models using age/gender dummies plus lagged spending by type of service to measure consumer expectations for each type of service[8].

B. *Selection Indices by Type of Service and Risk Adjustment*

Table 2 summarizes the selection indices for 33 type-of-service categories when pooling across the four plan types[9] and services are sorted by descending order of selection indices with no risk adjustment. The left three columns show the predictability,

---

[7] We observe that models using commercial data tend to have higher R^2 than those reported by EM using Medicare data. A similar finding was reported in Ash et al. (2000).

[8] We dropped three types of services in the TOS model results. We dropped "Dialysis" since there is a special program in Medicare for people with kidney function failure, hence spending on this type of service is incomplete. We dropped "Indian/tribal facility, no procedure code", due to its low mean. We also dropped "Missing, unknown". For each of the decompositions, we delete any service in which mean spending in any one of the four plan types is less than $0.50 per person per year.

[9] Table 2 includes individuals who switch from one plan type to another from 2003 to 2004 (assigning them their end of 2004 plan). Sensitivity analysis is done to check whether it matters if we drop enrollees who switch plan types. The results are highly robust to including or excluding this group, so we left them in for our primary analysis.

predictiveness and selection indices, assuming the same premiums for all plan types. The fourth and fifth columns list indices with risk-adjusted premiums by age-sex and by age-sex plus diagnoses respectively.

As discussed in our model section, the EM selection index, the product of the predictiveness and predictability, measures a health plan's incentive to ration medical service. Some services that are highly predictable (e.g. Home/Mobile clinic) are not particularly predictive and vice versa (e.g. Anesthesia). A larger magnitude of the index implies greater incentive for health plans not to provide such kind of service, so that services that are under-provided have larger index, e.g. home/mobile clinic, skilled nursing facility/intermediate care, home visits, hospice care etc. On the other end, ambulatory procedures, echography, major procedure, endoscopy, office visits, specialist, anesthesia, emergency room and standard imaging are the least rationed services. The selection index for predicted total spending captures the average incentive to select across all kinds of services and provides a useful benchmark for dividing services that tend to be over or under-provided.

The indices we obtain here are consistent with EM in terms of the order, but are larger in magnitude. The correlation between our indices and EM indices from Medicare on the same set of services is 0.63. One possible explanation for the difference is that the data used here comes from commercial plans, which are more profit-oriented than Medicare. We expect tighter rationing in commercial plans instead of in Medicare program that is funded by the government.

Table 2, also shows that the rankings of selection indices are robust to the level of risk adjustment. The magnitude of selection indices declines by 4 percent on average when

controlling only age and gender, and declines by 56 percent when additionally controlling diagnoses.[10]

Risk adjustment reduces, but does not completely eliminate, adverse selection incentives. The magnitudes of indices are highly correlated among three sets of indices with all correlation coefficients higher than 0.97. For the rest of the results, we focus on indices calculated without risk adjusted premiums. Results with risk adjustment are available upon request.

C. *Selection Effects by Plan Type*

We next compare selection indices among four plan types. We run separate regression models for each of the 33 TOS for each of the four plan types (132 TOS spending regressions in total). Predicted values from each model are then used to calculate selection indices separately for each plan type. Strikingly, as shown in Table 3, despite some differences in magnitudes, the incentives to select which services to ration are similar across all four plan types. For instance, home/mobile clinic, home visit and hospice are highest ranked services for most plan types. The correlations of indices between each individual plan and all plan types are greater than 0.86 for all plan types. And the correlations of rankings are greater than 0.93 in all cases. To demonstrate that the results are robust to risk adjustment, we repeat the same analysis as we have done in the full sample case. The same defined correlation of indices or rankings are all greater than 0.82. The results clearly show that managed and non-managed care plans have similar incentives to ration services. What seems to explain the

---

[10]Recall from equations 1 and 2 that as risk adjustment changes, so does the variability of profit. Without taking into account this change, risk adjustment reduces the selection index by 52 percent. Diagnostic risk adjustment explains 17% of the variance in total spending (Table A1), so the standard deviation with risk adjustment σ_{Π}^{RA} is about $\sqrt{1-0.17}$ of the standard deviation without risk adjustment σ_{Π}^{No}. Since this effect does not vary across difference service, the true selection indices with risk adjustment are all reduced further by a fixed percentage based on the results shown in Table 2. Combining the two changes, the magnitude of selection indices decline by at least 11 percent($\approx 1-0.98 * \sqrt{0.83}$) from age and gender risk adjustment and by 56 percent ($\approx 1-0.48 * \sqrt{0.83}$) from conventional diagnosis-based risk adjustment.

difference in enrollments and spending are that managed care plans are better able to moderate access or spending selectively on certain services, which affects both costs and enrollments.

D. *Selection Incentives versus Service Spending*

So far we have confirmed that selection incentives are robust to alternative econometric specifications and consistent across plan types. We next turn to whether services predicted to have the strongest selection incentive are in fact supplied least generously by plans that typically receive fixed premiums, particularly HMOs and POS plans. Figure 1 presents our key results graphically. The horizontal axis in the diagram is the selection index using the full sample. The vertical axis measures the MEANRATIO defined in section III. By construction, the weighted average of MEANRATIO for each type of service is one. Figure 1 shows that for services with low selection indices, HMO and POS plans are more likely to have MEANRATIOs above one while the COMP and PPO plans are more likely to have MEANRATIOs below one. The opposite pattern shows up for services with high indices. Only COMP has an upward sloping trend line. These results strongly suggest that HMOs, which usually receive fixed capitated premiums, are the most successful at achieving favorable selection and COMP plans (which are often experience rated and hence less concerned about costs) experience the worst adverse selection. The patterns of these figures are robust even if we remove the outliers, shown in Figure 2. These strong relationships between plan type and the correlation between spending by type of service and our service selection indices are highly statistically significant (Table 4).

One interesting observation is that spending on office visits and emergency room are the two services generously supplied by HMOs relative to non-managed care plans. This finding is consistent with Deb et al. (2006) who use Medical Expenditure Panel Survey (MEPS) data to show that, after accounting for self-selection, individuals enrolled in HMO

18

plans have 2 more visits to doctors and 0.1 more visits to emergency rooms than non-managed care plans per year.

In summary, all plan types have similar incentives to ration services to maximize profits as calculated by our selection indices. Although the incentives are similar, plan types differ in how effectively they respond to these incentives. COMP plans do not engage in selective contracting, nor typically have strong utilization review ability to influence spending by type of service. Moreover, many COMP plans use either third-party administration contracts or experience-rated premiums, so there is less incentive to control costs. Selective contracting, utilization review, and fixed premiums are much more common for HMO and POS plans, and our results suggest that, on average, these managed care plans are more successful at restricting services that have larger selection indices. Among four plan types, COMP plan has the worst selection of enrollees. Its relative spending on services goes up noticeably as selection indices increase, in comparison to the dropping MEANRATIO of other plan types.

V.   Conclusion

This paper applies the methodology proposed by EM (2007) to a large rich dataset, Thomson-Reuters MarketScan commercially insured data, for the period 2003-2004. We confirm the EM finding that selection incentives are strong for certain services commonly thought to be provided more by non-managed care than managed care commercial plans. We find that the EM selection indices are very similar across all four plan types even after risk-adjusted plan premium. By decomposing spending by type of service, we find that home/mobile clinic, skilled nursing facility/intermediate care, home visits, hospice, oncology, hospital visit, and durable medical equipment are most under-provided by managed care plans.

We find a compelling relationship between service-level spending and selection indices by type of service. Services predicted to be more tightly rationed are indeed found to comprise a smaller percent of total spending in managed care plans relative to non-managed care plans. Various specification checks verify that our results are robust to different information sets, different subsamples, different plan types, different risk adjustment models and different econometric specifications.

One important weakness of our analysis is that we are unable to estimate, and hence to incorporate, the effects of variations in demand elasticity for each type of service. Available information suggests that most outpatient services have a higher elasticity than inpatient services, and preventive care services are more elastic than acute care services. The EM formula implies that more elastic services will be subject to greater attention for stronger selection incentives than less elastic services.

We are also aware that our selection index is only a partial measurement of selection strategies used by insurance plans. Other than service level distortion focused in this paper, private insurers could directly advertise their plans to the targeted population; or they select favorable enrollees through benefit plan design; or they dump those undesired potentially high cost individuals. Our index would not capture those regulated services. But the index still characterizes the subtle incentives to ration services by insurers.

Despite the caveats, this study can help improve health plan payment policy. Our findings show that careful risk adjustment reduces selection incentives meaningfully, by an estimated 56%, although the incentives to select are not completely eliminated. Supply-side cost sharing and regulations on plan behavior are still needed in order to mitigate selection incentives. The services identified as prone to be distorted are important for policy makers to monitor so as to neutralize commercial plans' incentives. The results have implications for

managed care regulation, capitation formula, employment based insurance, provider payment, and health system research.

REFERENCES

Ash, Arlene, Randall P. Ellis, and Gregory C. Pope, et al. 2000. "Using Diagnoses to Describe Populations And Predict Costs." *Health Care Financing Review*, 21: 7-28.

Cao, Zhun and Thomas G. McGuire. 2003. "Service-level Selection by HMOs in Medicare." *Journal of Health Economics*, 22: 915-931.

Deb, Partha.,Chenghui Li, Pravin K. Trivedi and David M. Zimmer. 2006. "The Effect of Managed Care on Use of Health Care Services: Results from Two Contemporaneous Household Surveys." *Health Economics*, 15(7): 743-760.

Duarte, Fabian. 2011. "Price Elasticity of Expenditure across Health Care Services" The Selected Works of Fabian Duarte. Available at: http://works.bepress.com/fabian_duarte/1

Dusheiko, Mark, Hugh Gravelle, Steve Martin, Nigel Rice and Peter Smith.2009. "Modeling Individual Patient Hospital Expenditure for General Practice Budgets." unpublished.

Eggleston, Karen and Bir, Anupa. 2007. "Measuring Selection Incentives in Managed Care: Evidence From The Massachusetts State Employee Insurance Program." *Journal of Risk and Insurance*, 76(1): 159-175.

Ellis, Randall P., and Thomas G. McGuire. 2007. "Predictability and Predictiveness in Health Care Spending." *Journal of Health Economics*, 26 : 25-48.

Frank, Richard., Jacob Glazer and Thomas G. McGuire. 2000. "Measuring Adverse Selection in Managed Health Care." *Journal of Health Economics*,19: 829-854.

Glazer, Jacob and Thomas G. McGuire. 2000. "Optimal Risk Adjustment in Markets with Adverse Selection: an Application to Managed Health Care." *American Economic Review*, 90 (4): 1055-1071.

Glazer, Jacob, Thomas G. McGuire, Zhun Cao and Alan Zaslavsky. 2008. "Using Global Ratings of Health Plans to Improve the Quality of Health Care." *Journal of Health Economics*, 27: 1182-1195.

Manning, Willard G., Joseph P. Newhouse, Naihua Duan, Emmett B. Keeler, and Arleen Leibowitz, 1987. "Health Insurance and the Demand for Medical Care: Evidence from A Randomized Experiment." *American Economic Review*, 77(3): 251-277.

Meyerhoefer, Chad D., and Samuel H. Zuvekas. 2009. "New Estimates of the Demand For Physical and Mental Health Treatment", *Health Economics*, 19(3): 297-315.

Ringel, Jeanne S., Susan D. Hosek, Ben A. Vollaard, and Sergej Mahnovski., 2005. "*The Elasticity of Demand for Health Care: A Review of the Literature and Its Application to The Military Health System*", National Defense Research Institute, RAND Health.

TABLE 1---BASIC STATISTICS FOR FOUR PLAN TYPES

Plan type	N	Age in 2003	Total spending in 2003	Total spending in 2004
Comprehensive	555,087	48.9 (11.2)	4,868 (13,153)	5,879 (19,276)
PPO	1,580,272	44.6 (11.5)	4,258 (21,348)	4,838 (17,630)
POS	500,188	43.6 (11.0)	3,559 (10,367)	4,289 (15,436)
HMO	192,744	40.9 (10.6)	3,118 (8,923)	3,754 (12,509)
Full sample	3,285,751	44.9 (11.4)	4,120 (16,940)	4,788 (17,199)

Notes: Table reports means and standard deviations(in parentheses). For each individual plan, we only include adults (age>=21, <=64) who stay in the same plan type for 12 months in 2003 and do not switch from 2003 to 2004. For full sample, we add in switchers.

TABLE 2---PREDICTABILITY AND PREDICTIVENSS OF SPENDING BY TYPE OF SERVICE USING WEIGHTED LEAST SQUARES METHOD (N=3,285,751)

Spending by type of service categories	No risk adjustment			Selection indices with risk adjustment	
	Predictability CV($\hat{m}_s$)	Predictiveness Corr($\hat{m}_s$, M)	Selection indices	Premium risk adjusted by agesex	Premium risk adjusted by agesex+HCC
Panel A: Indices					
Home/mobile clinic	33.374	0.074	2.483	2.433	1.613
SNF/intermediate care[§]	27.345	0.083	2.274	2.192	0.930
Home visit	19.185	0.112	2.147	2.104	1.170
Hospice	20.919	0.101	2.119	2.023	0.789
Other procedure (inc unclassified)	7.421	0.257	1.910	1.870	1.033
Oncology	3.275	0.273	0.895	0.813	0.325
Hospital visit	2.365	0.374	0.886	0.832	0.309
Durable medical equipment	3.690	0.217	0.800	0.759	0.319
Office/clinic/amb center	2.266	0.339	0.768	0.724	0.327
Inpatient facility	1.752	0.406	0.712	0.656	0.247
2004 annual spending	**1.396**	**0.451**	**0.629**	**0.584**	**0.257**
Imaging/procedure	1.625	0.348	0.565	0.506	0.184
Major proc cardiovascular	1.818	0.307	0.559	0.476	0.168
Pharmacy	1.765	0.312	0.551	0.498	0.246
Advanced imaging:CAT[*]	1.994	0.270	0.537	0.500	0.178
Eye procedures	2.849	0.140	0.399	0.301	0.090
Minor procedures	1.523	0.254	0.387	0.356	0.138
Major proc orthopedic	1.428	0.264	0.378	0.319	0.100
Lab tests	1.025	0.357	0.366	0.334	0.127
Consultations	0.902	0.385	0.347	0.309	0.101
Advanced imaging: MRI[¶]	1.197	0.282	0.338	0.309	0.099
Standard imaging	0.922	0.324	0.299	0.239	0.078
Emergency room	1.327	0.217	0.288	0.289	0.089
Anesthesia	0.800	0.352	0.282	0.253	0.081
Specialist	3.110	0.090	0.280	0.276	0.081
Other tests	0.924	0.298	0.276	0.235	0.068
Office visits	0.765	0.306	0.234	0.207	0.064
Endoscopy	0.776	0.299	0.232	0.184	0.053
Major procedure	1.122	0.206	0.231	0.234	0.073
Echography	0.970	0.233	0.226	0.202	0.057
Ambulatory procedures	0.791	0.269	0.213	0.191	0.054
Panel B: Correlation between indices					
No risk adjustment			1.00	1.00	0.97
Premium risk adjusted by Agesex				1.00	0.97
Premium risk adjusted by agesex+HCC					1.00

Note: Our estimation sample includes adults(>=21, <=64) from all plan types, who are eligible for all of 2003 and for any part of 2004. We include switchers between two years in our sample. Regression uses weighted LS and takes fraction of year eligible as sample weights. Type of service categories are aggregates of the Berenson-Eggers Values. Selection indices are reported as the product of two terms, predictibability and predictiveness and are sorted by the decending order of selection indices without risk adjustment.

[§]SNF: Skilled nursing facility. [*]CAT: Computerized axial tomography. [¶]MRI: Magnetic resonance imaging.

TABLE 3---THE MAGNITUDE AND RANKING OF INDICES OF EACH TYPE OF SERVICE
FOR 4 PLAN TYPES

Spending by type of service categories	Selection indices					Rankings of selection indices				
	All plan types	COMP	PPO	POS	HMO	All plan types	COMP	PPO	POS	HMO
Part A: Selection indices and rankings										
Home/mobile clinic	2.483	2.866	1.364	6.578	1.935	1	1	5	1	4
Home visit	2.274	1.644	2.938	1.962	1.097	2	5	1	5	7
Hospice	2.147	1.779	1.993	2.162	2.026	3	4	3	4	3
SNF/intermediate care[§]	2.119	1.932	2.032	2.672	2.554	4	2	2	2	1
Other procedure (inc unclassified)	1.910	1.824	1.845	2.553	2.221	5	3	4	3	2
Oncology	0.895	0.899	0.846	1.048	1.223	6	7	6	7	5
Durable medical equipment	0.886	0.956	0.819	1.196	1.100	7	6	7	6	6
Office/clinic/amb center	0.800	0.887	0.704	1.027	0.794	8	8	9	8	10
Hospital visit	0.768	0.783	0.749	0.706	0.959	9	9	8	12	8
Inpatient facility, no proc	0.712	0.735	0.660	0.922	0.796	10	10	10	9	9
2004 annual spending	**0.629**	**0.654**	**0.607**	**0.714**	**0.635**	**11**	**11**	**11**	**11**	**13**
Pharmacy	0.565	0.653	0.515	0.720	0.656	12	12	14	10	12
Major proc cardiovascular	0.559	0.533	0.531	0.698	0.680	13	14	13	13	11
Imaging/procedure	0.551	0.512	0.558	0.548	0.520	14	15	12	15	14
Advanced imaging:CAT[*]	0.537	0.617	0.505	0.624	0.509	15	13	15	14	15
Eye procedures	0.399	0.371	0.397	0.420	0.355	16	22	16	17	20
Major proc orthopedic	0.387	0.396	0.355	0.432	0.493	17	20	18	16	16
Minor procedures	0.378	0.410	0.345	0.400	0.401	18	18	19	19	18
Lab tests	0.366	0.450	0.355	0.417	0.390	19	16	17	18	19
Advanced imaging: MRI[¶]	0.347	0.434	0.337	0.369	0.297	20	17	20	20	24
Consultations	0.338	0.384	0.321	0.361	0.343	21	21	21	21	21
Standard imaging	0.299	0.295	0.290	0.328	0.308	22	25	22	24	23
Anesthesia	0.288	0.397	0.282	0.353	0.221	23	19	23	22	28
Ambulatory procedures	0.282	0.316	0.275	0.302	0.269	24	23	25	25	26
Emergency room	0.280	0.259	0.271	0.329	0.261	25	28	26	23	27
Major procedure	0.276	0.294	0.275	0.298	0.270	26	26	24	26	25
Specialist	0.234	0.252	0.233	0.236	0.219	27	29	27	31	29
Other tests	0.232	0.242	0.220	0.245	0.435	28	31	29	29	17
Endoscopy	0.231	0.304	0.212	0.256	0.319	29	24	30	27	22
Echography	0.226	0.274	0.224	0.239	0.203	30	27	28	30	31
Office visits	0.213	0.245	0.198	0.246	0.211	31	30	31	28	30
Sample size	3,285,751	555,087	1,580,272	500,188	192,744					
Part B. Correlation with indices or rankings of all plan types										
		0.973	0.938	0.864	0.915		0.965	0.990	0.984	0.937

Note: Our estimation sample includes adults(>=21, <=64) from all plan types, who are eligible for all of 2003 and for any part of 2004. We include switchers in all plan types sample but only nonswitchers for each individual plan type. Regression uses weighted LS and takes fraction of year eligible as sample weights. Type of service categories are aggregates of the Berenson-Eggers Values. Selection indices are reported as the product of two terms, predictibability and predictiveness without any risk adjustment on premiums and are sorted by the decending order of selection indices by all plan types.

[§]SNF: Skilled nursing facility. [*]CAT: Computerized axial tomography. [¶]MRI: Magnetic resonance imaging.

TABLE 4---CORRELATIONS COEFFICIENT BETWEEN
INDICES AND MEANRATIOS BY TYPE OF SERVICE

CORR(. , .)		Meanratio			
		COMP	PPO	POS	HMO
Selection indices	All plan types	0.689*** (0.137)	-0.213 (0.185)	-0.583*** (0.154)	-0.474*** (0.166)
	COMP	0.690*** (0.137)	-0.179 (0.186)	-0.626*** (0.147)	-0.452*** (0.169)
	PPO	0.662*** (0.142)	-0.263 (0.182)	-0.482*** (0.166)	-0.468*** (0.167)
	POS	0.711*** (0.133)	-0.209 (0.185)	-0.642*** (0.145)	-0.438*** (0.170)
	HMO	0.698*** (0.135)	-0.160 (0.187)	-0.714*** (0.132)	-0.435*** (0.170)

Note: Table reports the correlation between selection indices and meanratios. The indices are calculated without risk adjustment on premiums. We include switchers in all plan types sample but only nonswitchers for each individual plan type.

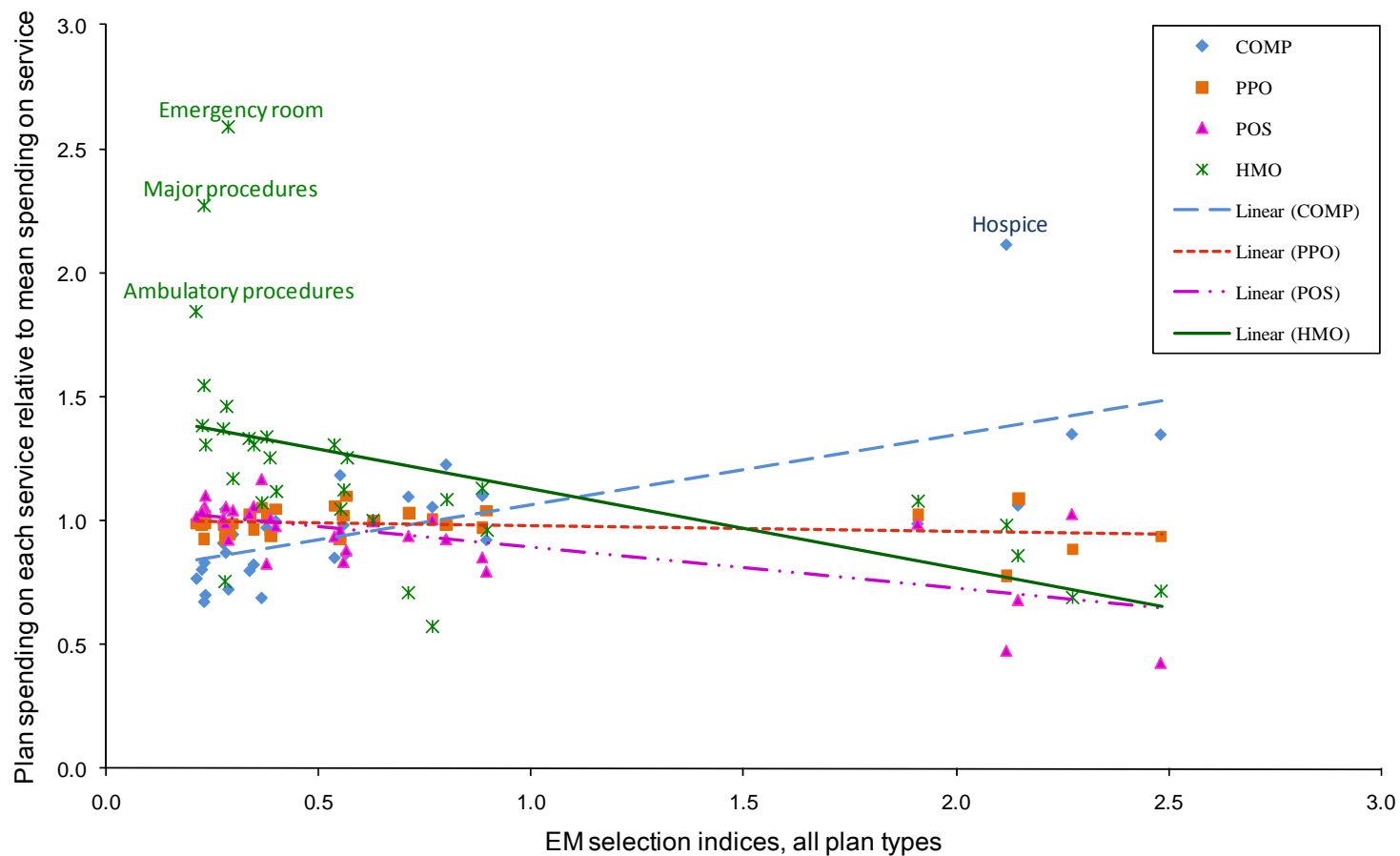***: 1%, **:5% and *:10% of significance.

FIGURE 1 PLOT OF SELECTION INDICES VERSUS RELATIVE SPENDING BY PLAN TYPE
(TYPE OF SERVICE)

Note: This figure is a plot of plan percentage spending on each service relative to the mean percentage spending on that service by selection calculated using the pooled sample of all plan types. Service is decomposed by type of service.
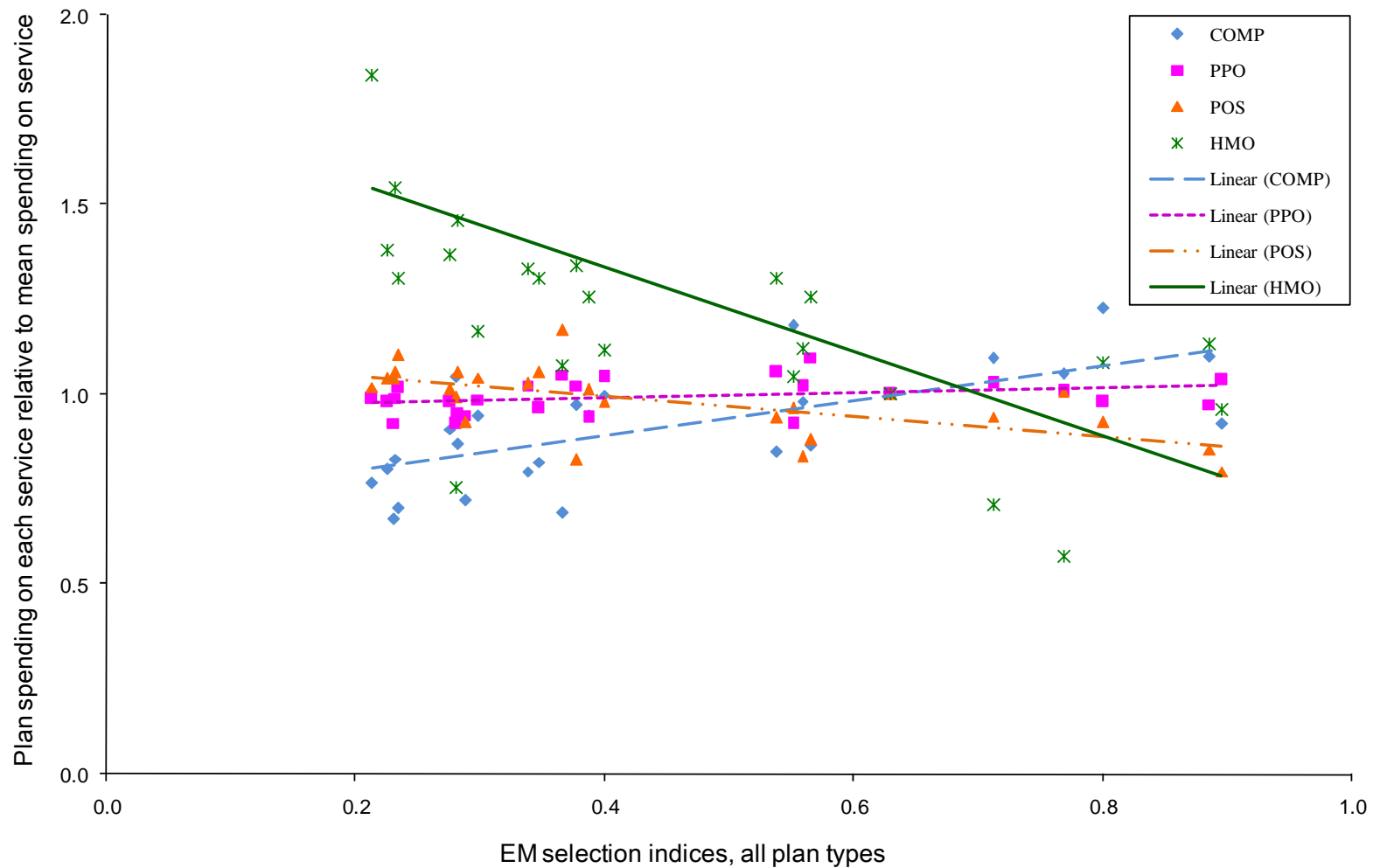
FIGURE 2 PLOT OF SELECTION INDICES VERSUS RELATIVE SPENDING BY PLAN TYPE OMITTING ALL
SERVICES WITH SELECTION INDICES GREATER THAN 1.0
(TYPE OF SERVICE)

Note: This figure is a plot of plan percentage spending on each service relative to the mean percentage spending on that service by selection indices of all plan type. We omit services with indices larger than 1. Service spending is decomposed by type of service.

Appendix A

In this section we reproduce derivation of the selection index formula proposed by Ellis and McGuire (2007).

A health plan offers S services. An individual's expectation about the value of services she will receive determines her choice of the plan. Let $\hat{m}_{is}$ denote the amount that individual i expects the plan will spend on providing service s and let $\hat{m}_i = \{\hat{m}_{i1}, \hat{m}_{i2}, ..., \hat{m}_{iS}\}$. The benefit to individual i from a plan is $u_i(\hat{m}_i) = v_i(\hat{m}_i) + \mu_i$, where μ_i is a random term with distribution function Φ_i. Utility from services is assumed to be additively separable, and hence it can be written as

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

Individual i chooses this plan if

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

where $\bar{\mu}_i$ is the valuation the individual places on the next best alternative plan.

A managed care plan is assumed to efficiently ration the amount of health care each patient receives. Let q_s denote the service-specific shadow price a plan sets for service s. A patient will receive a quantity of services, m_{is}, which is determined by:

$$v_{is}'(m_{is}) = q_s, \text{ where } m_{is} \text{ is actual spending on services } s.$$

The plan chooses a vector of shadow prices $q = \{q_1, q_2, \cdots, q_S\}$ to maximize its profits

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

where r is the revenue the plan receives for each individual and $n_i(\hat{m}_i) = 1 - \Phi_i(\bar{\mu}_i - v_i(\hat{m}_i))$ is the probability that health plan expects individual i would choose the plan. All individuals are assumed to share the same elasticity of demand for any service but elasticities can differ across services. Also the enrollment function is assumed to be uniform and the same for all i, i.e. $\Phi_i' = \phi$.

The derivative of profit with respect to $\{q_1, q_2, \cdots, q_S\}$ is

$$\frac{\partial \pi}{\partial q_s} = \eta_s q_s \sum_i \phi \hat{m}_{is} \pi_i - \eta_s \sum_i n_i m_{is}$$

(see FGM 2000). If we assume that plans have rationed services optimally, then at the first best $q_s = 1$. Let $\bar{m}_s$ be the average spending on service s. The profit incentive to raise q_s for service s is the index I_s, defined as $I_s \equiv \frac{\partial \pi}{\partial q_s} * \frac{1}{\bar{m}_s}$. I_s quantifies the profit incentives to ration service s, normalized by average expected spending. Therefore we have

$$I_s = \eta_s \left(\phi \sum_i \frac{\hat{m}_{is} \pi_i}{\bar{m}_s} - 1 \right), \text{ since } \rho_{\hat{m}_s, \pi} = \left(\frac{\sum_i \hat{m}_{is} \pi_i - \bar{m}_s \bar{\pi}}{\sigma_{\hat{m}_s} \sigma_\pi} \right) \text{ and } \rho_{\hat{m}_s, \pi} = -\rho_{\hat{m}_s, M} \text{ in the absence of any}$$

risk adjustment, we rewrite the selection index as $I_s = \phi \cdot \sigma_\Pi \eta_s \left(\frac{\sigma_{\hat{m}_s}}{\bar{m}_s} \rho_{\hat{m}_s, \Pi} - C \right), \quad C = \bar{\pi} - \frac{1}{\phi}$

TABLE A1---PREDICTIVE POWER OF VARIOUS INFORMATION SETS AND VARIOUS MODELS

	Weighted LS	OLS	Two-part Linear model
Part A			
Partial year eligibles included?	Yes	No	No
Sample mean	4500	4365	4365
Number of observations	3,285,751	2,894,687	2,894,687
Part B: R^2			
Age and gender only	0.016	0.018	0.018
Prior year total spending	0.087	0.090	0.094
Diagnosis organized by DCG/HCC	0.167	0.173	0.173
Covered charges by type of service	0.203	0.218	0.219

Note: The dependent variable is 2004 total covered charges. Our sample includes adults($>=21$, $<=64$) who stay in the same plans for 12 months in 2003. Partial enrollees in 2004 are only included in weighted LS regression and fraction of year eligible is used as sample weights. All regressions include 10 age-gender dummy variables.