# Variation in pricing, treatment, coding intensity and patient severity in spells of health care treatment

Randall P. Ellis and Wenjia Zhu

Boston University

June 8, 2014

Acknowledgements. We are grateful to Arlene Ash, Francesco Decarolis, Tim Layton, and Tom McGuire for useful comments and insights, although the authors bear full responsibility for any errors or omissions. This research was partially supported by Verisk Health through a grant to Boston University. The MarketScan Commercial Claims and Encounter data used for this project was licensed from Truven Analytics.

Keywords: health care spending, treatment spells, risk adjustment, exclusive provider organizations, consumer driven health plans

JEL codes:

# Variation in pricing, treatment, coding intensity and patient severity in spells of health care treatment

Abstract: This paper examines variation in health care utilization and spending of insured employees at large firms in order to understand the effects of three new health plan types that are growing in popularity: exclusive provider organizations (EPOs) which narrow the panel of available providers, and consumer-driven health plans (CDHPs) and high-deductible health plans (HDHPs) which raise cost sharing and promote wider consumer choice. We model decisions at the level of "treatment spells" which are fixed length periods of 30 (or 31) days that commence with a service after a gap in provider contact. The large apparent savings of EPOs, CDHPs and HDHPs are greatly reduced after risk adjustment, and controlling for endogenous plan selection, employer and consumer fixed effects, suggesting that their apparent savings are due to selection. None of the plans particularly promotes using lower price providers, and coding intensity does not vary meaningfully across plan types. Two stage least squares estimates find more significant effects on visits, relative to payment decisions. We find no evidence of significant savings from EPOs in this sample.

## Section 1: Introduction

There is enormous interest in understanding how to control health care costs, with many innovations in demand-side cost sharing, provider network design, and provider payment reforms being introduced and evaluated. Almost all analyses of these innovations use either annual spending per year or spending per visit (or even per claim) as their metric of costs, even though neither a year nor a visit is an ideal unit of decision making. Monthly and quarterly spending and utilization are increasingly being analyzed, but fixed calendar intervals chop spells of treatment into arbitrary chunks of time that do not usually correspond to decision making units. Episodes of treatment are an alternative unit of decision-making, but the challenges of classifying treatment into distinct episodes in the presence of the practically infinite set of combinations of conditions that are common among the chronically ill has hindered the adoption of this approach for episodes not centered around inpatient treatment. In this study we analyze decision-making at the level of spells of treatment that lie intermediate between monthly and episodic units of analysis and provide a potentially useful framework for modeling cost and utilization variation. We illustrate the usefulness of our approach by studying the effects of three recent health plan design innovations on patient and provider treatment decisions.

Consumer and provider decisions, which jointly determine treatment costs and outcomes, are influenced by both supply- and demand-side incentives. While both types of incentives are being changed, we still do not really know which approach is more effective at reducing costs, and if so, *how* cost reductions are achieved. Provider-level studies consistently find that providers vary significantly in their costs of treating a given condition (e.g., a sore shoulder), but it is not well understood how to attain cost reductions, and if a health plan type does reduce costs, whether the reductions occur due to lower prices, lower treatment intensity for that condition, or lower effort at identifying diseases to treat (or at least code) for a patient arriving with a general symptom. Similarly, consumer-level studies find suggest that restricting patient choice or raising cost sharing will reduce spending due to the moral hazard effect, but it is unclear whether any reductions occur because patients seek lower price providers, accept lower intensity treatment options, or choose providers that are less aggressive about identifying or coding diseases.

This paper incorporates state-of-the-art risk adjustment tools to control for variation in the underlying illness severity of the patients in order to evaluate their decision making on the various components that collectively affect total health payments. The decisions modeled include whether health plans incentivize patients to make fewer new or continuation visits, choose lower price providers, switch to a new PCP, accept lower treatment intensity when ill, or encourage providers to identify fewer or less serious conditions to treat. While there is research on each of these topics, we are not aware of any paper that has examined them in a large sample of patients with a unified analytic approach.

An important way of learning about consumer and provider treatment choices is to observe how treatment decisions change when consumers are subject to new incentives. The setting that we study is changes in the types of health plans offered in a sample of 66 large employers over five years in the US. Three common plan types in our sample represent the majority of health plan enrollment in the US: preferred provider organizations (PPOs), health maintenance organizations (HMOs) and point-of-service plans (POS). A fourth plan type, the comprehensive (COMP) plans, is being phased-out by most employers. Three health plan innovations are of particular interest since they are growing in popularity, and represent dramatically different approaches to cost containment. High-deductible health plans (HDHPs) impose high deductibles that increase demand-side cost sharing while emphasizing choice and promoting use of an employee's own health reimbursement account. Consumer-driven (or consumer-directed) health plans (CDHPs) are similar, but soften high deductibles while offering employer-sponsored health reimbursement arrangements. A different approach is taken by exclusive provider organization (EPO) plans that greatly restrict the choice set of providers that are covered, while reducing demand-side cost sharing for the covered providers. (Appendix table A-1 contains more detailed plan type definitions.) The central issue that this paper models is how health treatment decisions change when consumers change to a new health plan type, and whether plans types like EPOs and HMOs that restrict consumer choice while reducing cost sharing have a different effect than plan types like HDHP and CDHP that broaden consumer choice while increasing consumer cost sharing incentives.

Understanding how health plan designs influence spending has taken on increased importance in light of the significant changes implemented by the Affordable Care Act (ACA). Cost sharing plays a big role in the new Health Insurance Marketplace (HIM) with its offerings of bronze, silver, gold and platinum plans. Narrow panels of providers, and in particular EPOs, also figure prominently in the HIM, despite the lack of substantial evidence about their effect or popularity. Using risk adjustment to reallocate funding and quantify performance is another key element of the ACA. This paper attempts to contribute to our understanding of how health plans that differ in their cost sharing and breadth of provider choice affect health care spending, while using diagnosis-based risk adjustment and patient fixed effects to control for patient severity and taste variation.

Four features distinguish our approach from the existing literature: we focus on a unit of analysis called treatment spells; we decompose total spending into five components; we use rich and original risk adjustment models; and we use big datasets that enable us to identify nuanced choices with greater power. Econometrically, we use patient and provider fixed effects to control for time invariant taste, environmental and practice style factors, risk adjustment to control for time-varying illness patterns, and instrumental variables to control for endogenous health plan choices.

While the year as a unit of observation is useful for studying how broad patient characteristics affect spending and utilization, a year is too coarse for examining most consumer and provider

decisions: a given patient will often see many providers and have multiple spells of treatment in a given year. Keeler and Rolph (1988) was the first careful study of health care spending and utilization within a year, and developed a rich framework at the level of episodes of health care. Episodes of treatment start after a period in which no health care is received, continue while that condition is being treated, and end once no further visits are observed for a period of time. While attractive conceptually, remarkably few academic studies have used episodes for studying both inpatient and outpatient utilization decisions. Problems include distinguishing multiple, overlapping episodes, deciding when information becomes known and when continuation decisions are made, dealing with incomplete episodes at the start and end of the sample period, and incorporating chronic illnesses that may persist indefinitely. In practice, episode algorithms assign a substantial fraction of total spending either to incomplete episodes or to chronic condition episodes that last for the entire calendar year, with both requiring distinct models.

A second and related approach is to use monthly spending and utilization measures. Manning et al (1987) and Ellis (1986) were early users of this approach, with more recent examples being Einav et al. (2013), and Finkelstein (2014). All of these studies use a monthly unit of observation in order to understand the effects of nonlinear insurance budget constraints on health care utilization, using variation in health care cost sharing within a year to study demand effects. Similar to episodes, monthly models are challenged in capturing multiple conditions and dealing with end-of-sample effects. None of the existing papers use careful risk adjustment models to distinguish health care status, or have used employer, patient and provider fixed effects.

The treatment spell approach we adopt here lies intermediate between episodes and monthly analysis, and has its origins in the early work of Manning and Keeler in modeling mental health spells of treatment (Manning et al 1988). Similarly to episodes, we time the beginning of a new treatment spell as of a new visit after a period with no treatment, but rather than ending when treatment is completed, we examine visits and spending for a fixed time interval of 30 (or 31) days. The advantage of treatment spells over monthly use is that they capture initial use more uniformly, since calendar months can capture anywhere from one to 31 days of treatment. Unlike Manning et al, we do not attempt to uniquely assign utilization and spending variables to distinct illnesses, but instead use a risk adjustment framework to capture the full set of diagnoses that may be present during a given monthly spell. Hence we do not classify a spell as a diabetic, hypertensive or mental health spell, but rather analyze the spell as being by someone with all three conditions. Similarly to a simple monthly model, we can distinguish “new visit” spells from “continuation” spells, although rather than restarting on the first day of the calendar month, our continuation spells start immediately subsequent to a previous new or continuation treatment spell. Fitting around fixed length treatment spells are “no visit” spells of variable length, which are simple to model since by construction there is no spending or visits in them. Figure 1 provides an example of how visits by one person can be grouped into new and continuing spells of treatment, as well as into “no treatment” spells that lie in between them. Existing risk adjustment approaches are readily adapted to predict spending and use of services in fixed-

duration spells rather than variable-length episodes or months with variable number of days of treatment.

At the heart of our analysis is a decision to decompose health care spending over a fixed period of time into five components. Following Manning et al (1987) health care spending can be thought of as the product of the probability of any treatment and expected spending when positive. Conditional on spending being positive, spending can be further decomposed into the product of four further terms: pricing intensity, treatment intensity, coding intensity, and patient severity of illness. We take patient severity of illness as exogenous, but estimate separate models of visits, pricing, treatment and coding intensity to understand how different insurance plans types affect each of these health care decisions.

Our study builds upon a vast literature. Ever since Wennberg et al (1973), practice-level variation has been a central area of research, with much of the research focusing on treatment or total cost variation, often without trying to distinguish pricing, quantity, disease discovery, or patient severity variation. Following recent research that highlights the central role of primary care practitioners (PCPs) in their diagnoses, treatment and referral decisions (e.g., Berenson, et al. 2010), we study provider variation by assigning visits and costs to PCPs, and not to specialists or facilities, assuming that PCPs influence the discretionary use of labs, hospitals and specialists through their referral patterns (Robertson, 2011).[1]

Also relevant is the growing literature that finds significant consumer heterogeneity in health treatment decisions, made possible in part by large datasets that enable detection of second order effects of plan features. Cutler, Finkelstein and McGarry (2008) document risk aversion heterogeneity, Einav et al. (2013) document heterogeneity in demand responsiveness, and Cutler et al (2014) characterize heterogeneity in consumer preference using "strategic surveys". A separate strand of literature, exemplified by Dafny (2013) models heterogeneity in preferences for health insurance. We capture patient heterogeneity in three ways, via the worker's employer, via time-invariant patient fixed effects which capture taste variation, and via time varying disease burdens as captured by time-varying patient risk scores. Ellis (2012) argues that variation in health status should not be interpreted as variation in tastes, and notes that studies that do not adequately control for health status will attribute too much variation in decisions to consumer taste heterogeneity rather than severity of illness variation.

An innovation of our paper is that we attempt to distinguish severity of illness from coding intensity. Coding intensity variation and strategic diagnostic coding have recently received considerable attention, perhaps because of the increased use of risk adjustment in the ACA's new Health Insurance Marketplace (HIM) previously known as Health Insurance Exchanges. Recent

[1] We would love to assign treatment spells not only to PCPs, but also (nonexclusively) to other types of providers, however our sample size does not enable us to do this while assigning adequate number of spells to a given provider ID. We await the use of all-payer data that would enable us to take on this even more challenging objective.

articles by Wennberg and colleagues (2013, 2014) have challenged the Medicare risk adjustment formula by arguing that the Medicare formula unduly reflects variation in coding intensity as well as variation in severity (For a contrary view, see Newhouse et al (2013)). Geruso and Layton (2014) find evidence of significant coding intensity increases in response to risk-adjusted payments in private plans in Medicare. Without taking a position on whether a risk adjustment formula used for payment should or should not be adjusted for coding intensity variation, we estimate whether coding appears to be responsive to plan types within our sample.

To accommodate the new treatment spell approach we estimate risk adjustment models that predict spending and utilization at the monthly level. We use a "point-in-time prediction" framework that continually updates the information used to predict features of each 30-day spell using information from the twelve months prior to the start of that treatment spell. This approach is similar to the "continuous update" framework used in Ellis and Ash (1989), although they model calendar months rather than treatment spells. We use a combination of prospective and concurrent risk models and use different information sets to try to disentangle *ex ante* from *ex post* diagnostic coding for a treatment spell as a way of quantifying provider coding intensity.

Very large samples are needed in order to reliably calibrate rich risk adjustment models, and provider style can only be captured if a moderate number of observations are available on each health care provider. Precise estimates of treatment effects of new plan types also require very large samples given the high coefficient of variation in monthly health care spending (the standard deviation of spending across treatment spells is 5 times the mean). Therefore we use very large samples of insurance claims. Our initial sample contains 5.1 million commercially insured adults continuously eligible for five years, and this sample of over 300 million treatment spells is used to calibrate our various risk adjustment models. As outlined below, we take smaller subsets to analyze health plan effects, with our core analysis using 62 million treatment spells for which we can assign an employer, PCP, and plan type.

Strengths of our data are not only the large size but the diverse set of employers and plans offered; but these strengths come at the cost of imprecision about health plan features. We conduct our analysis focusing on results from 255 employer-years, offering an average of 3.07 plan types. A drawback of our data is that we do not have premium or benefit plan information for individual plans, and specific plan identifiers are missing in many years for most enrollees. We accommodate this by using employer-year-coverage type-plan type dummies, and instrument these dummies to control for their endogeneity.

A potential criticism of our work is that in our efforts to include all types of spending and all types of providers and health plans, we end up aggregating diverse treatment types, diseases and complex decisions. In our defense we note that countless other studies have examined the effects of individual, provider, and plan characteristics on aggregates such as total spending, inpatient treatment and plan choice using even more aggregated approaches than ours, such as annual or county level spending. In our comprehensive rather than selective approach, we find new

insights that we hope overcome the justifiable concerns about aggregation. Examination of disaggregated procedures and disease patterns, for which treatment spells are well suited, awaits further research.

The rest of the paper is organized as follows. Section 2 provides the conceptual framework underlying our empirical model. Section 3 introduces our data, variables of interest, and summarizes how we assigned patients to PCPs, while Section 4 discusses our econometric strategies for dealing with endogeneity and correlated errors. In Section 5, we present our descriptive analysis and regression results from various econometric specifications at diverse levels of unit of analysis. Section 6 concludes and briefly discusses the implications of this paper.

## Section 2: Conceptual Framework

### Section 2.1: Determination of health care use

We conceptualize health care use as determined jointly by both the consumer and the provider. Each year an employer chooses a set of health plans offered to employees. Meanwhile, PCPs decide which health plans to contract with and negotiate the fee structure they will receive from each plan for each service. Then employees choose from among the available health plans, with choices reflecting plan characteristics, employee demographics, employer, and household health status. Employee plan choice and PCP plan contract choices may or may not reflect careful optimization each year. Once each calendar month, each adult in the household decides whether to seek (or continue) treatment given their health plan, tastes, current and past health status, their PCP, and whether treatment was sought the previous month. Conditional on deciding to seek treatment ("make a visit"), and knowing not only the lagged characteristics of her own severity of illness but also the broad features of her current health status in the current month, the patient receives care from a set of diverse providers, who may or may not include their own previous PCPs. Influenced by their PCP's own specialty and practice setting, health plan features (including its negotiated prices), and the patient's characteristics, the PCP and the patient jointly decide on the services to provide or referrals to give to diverse laboratories, specialists, hospitals and emergency departments. Many services are also provided without any referrals.

We do not attempt to model consumer and provider decisions separately, since we only observe the joint outcome of both decision processes. For this paper, we do not separately model inpatient and outpatient treatment decisions, but capture this choice only in terms of the implied treatment intensity. Assigning treatment spells to PCPs enables us to analyze patterns at either the patient or the PCP levels, and importantly allows for more refined analysis of within-year patient and provider joint decisions.

### Section 2.2: Decomposition of Overall Payment Intensity

We decompose total spending on each monthly spell of treatment, where i indexes patients, j providers, s services, and t the month. We model total real spending (whether consumer paid or plan paid) on a monthly spell as the product of five terms, each of which is potentially influenced by a consumer's health plan.

$$
\begin{aligned}
&Expected\ total\ health\ spending\ \ of\ patient\ i\ with\ PCP\ j\ in\ a\ treatment\ spell\ t,\ Y_{ijt} \\
&= E[Y_{ijt}] = \left[\Pr(Y_{ijt} > 0)\right](Y_{ijt}|Y_{ijt} > 0) \\
&= \left[\Pr(Y_{ijt} > 0)\right]\left[\frac{\sum_s P_{ijst}Q_{ijst}}{\sum_s P_{i.st}Q_{ijst}}\right]\left[\frac{\sum_s P_{i.st}Q_{ijst}}{f(H_{ij.t})}\right]\left[\frac{f(H_{ij.t})}{g(H_{i..t-1})}\right][g(H_{i..t-1})] \\
&= (visit\ decision)(price\ intensity)(treatment\ intensity)(coding\ intensity)(severity) \quad (1)
\end{aligned}
$$

Equation (1) is an identity; each component in brackets can be calculated empirically, with the natural specification being in logs since the terms all enter multiplicatively. The first term in brackets is the visit decision, which is the probability of starting or continuing a spell of treatment in a given month, and thereby incurring positive spending. The second term is a conventional price index with weights being the spell's actual quantities, using the provider's actual prices in the numerator and national average prices in the denominator. The third term is a measure of treatment intensity. The numerator uses national average prices of each service to weight actual quantities provided during the spell. The denominator $f(H_{ij.t})$, is a measure of predicted spending given the diseases concurrently assigned by the patient i's own provider j for that month of treatment t. Overall this third expression reflects the amount of services provided to patient i during the spell given condition diagnosed by PCP j in that spell. Intuitively, this term is the quantity of services, as measured using the national procedure price average, relative to how sick the patients are as measured by coding by providers for that spell of treatment.

The fourth term in brackets is novel and therefore deserves more careful description. It is fundamentally the ratio of two risk adjustment predictions. The numerator of this term, matching the preceding denominator, is the expected cost given the provider's own concurrently assigned diagnoses during the current spell month of treatment. The denominator, as well as the final term in brackets, $g(H_{i..t-1,})$, is a different measure of expected spending, capturing the consumer's expected spending (or patient severity) at the time of making the decision to seek treatment. At a minimum, it should include the diseases known to the consumer from the prior twelve months, but the consumer may also know something about the nature of their current spell (e.g., an injured shoulder, sore throat, feeling of sadness). Consequently, the ratio in the fourth term captures variations across providers in diagnosing patients of the same general illness severity. In equation (1), we consider the two functions $f(H_t)$ and $g(H_{t-1})$ as risk adjustment predictions. There are many possible ways of calculating the two risk adjustment measures, which we now turn to.

### Section 2.3: Risk Adjustment

Risk adjustment models have gained increasing attention in the US in the past fifteen years with their expanded use for paying Medicare Advantage (MA), Medicaid managed care (MMC) plans, Medicare Part D, and Health Insurance Marketplace (HIM). The age, gender and diagnoses of an individual over any base period have been shown to be important predictors of many outcomes, notably including both concurrent models in which diagnoses are used to predict spending in the same base period as the diagnoses are observed, and prospective models in which spending and utilization are predicted using diagnoses from a previous base period (Ash et al, 2000).

We use diagnosis-based risk adjustment models in four ways in this paper: to capture patient illness burdens when modeling employee choice of health plans; to capture illness burden when consumer's decide to make a visit; to control for patient severity in treatment decisions; and to quantify coding intensity variation. Health plan type choice is found to be best predicted by a prospective model that predicts the sum of all adult family members spending. Spell month treatment decisions are modeled as depending on various combinations of concurrent and prior period diagnoses. Our method of capturing coding intensity is discussed below.

We use Verisk Health/DxCG Risk Solutions Version 4.21 software to generate predictions using eligibility, age, sex and diagnostic information. The DxCG models offer a choice of concurrent and prospective models as well as different levels of clinical coding details that can be used to examine whether consumer decisions are better captured by coarse (simplified) or fine (rich) distinctions within a disease category. For modeling health plan choice and visit decisions, we use a prospective, rich HCC classification system, DxCG's expansion of the HCCs used for the HIM. For overall severity (and also for the denominator of the coding intensity measure), we use the relatively coarse Related Condition Categories (RCCs) for prediction.[2] The 117 RCCs approximate the level of detail that most consumers will have, and are similar to conditions appearing on household surveys such as the Medical Expenditure Panel Survey (MEPS) administered. Intuitively, when choosing whether to make a visit or even how much to spend on care, the relevant information for the consumer is probably "shoulder disorders and injuries" rather than "bursitis of the shoulder" since the latter would presume a medical examination. In other words, the distinction between RCC and HCC classification can be thought of as a two-step process for capturing patient severity in that patients do not develop precise understanding of their conditions until seeing their doctors. Finally, we measure provider coding intensity as the

[2] The DxCG software allows models to be built using alternatively age and gender only, 31 Aggregated Condition Categories (ACCs), 117 Related Condition Categories (RCCs), 394 Hierarchical Condition Categories (HCCs) or 1113 Diagnostic Groups (DxGroups). As just one illustration, the ACCs would identify someone who has Diabetes, RCCs would distinguish Type I from Type II Diabetes, HCCs would distinguish whether the Type II diabetic has major or minor complications, and the DxGroup level categories would distinguish whether the major complications were for neurological or peripheral circulatory manifestations. Perhaps this example helps motivate our choice to use RCCs to capture consumer-understood severity for modeling decisions that reflect the expected level of spending. RCCs worked well in predicting annual spending in Australia (Ellis et al, 2013).

ratio of the HCC-based concurrent prediction (recalibrated via a new regression to predict the monthly spending) to the coarser RCC-based prediction, the idea being that this coarser system is less sensitive to coding intensity variation than the richer HCC system.

The risk scores used in modeling visit and treatment decisions are calculated using a "point-in-time" prediction strategy. Diagnostic information from the 12 months prior to the start of the treatment spell is used to predict resource utilization and costs in that month. For example, the consumer's severity (and thus expected spending) in month 1 of 2008, is predicted using the array of RCCs from 2007:1 to 2007:12 as well as the array of RCCs specifically for 2008:1. Information from 2007:2 through 2008:2 is then used to predict spending in 2008:2, and so on.

Our choice of the DxCG classification for our analysis is driven primarily by convenience and familiarity. A similar analysis could be done using either the coarser HCC classification currently used by the HIM, or other classification systems such as the Elixhauser Comorbidity Scores or The John's Hopkins Adjusted Clinical Group System. However these alternative systems do not generate concurrent and prospective relative risk scores (RRS) already calibrated to predict a single month of spending using a single month of diagnoses, which we will utilize to capture provider coding decisions. The predictive power of coarse versus rich sets of information, and prospective versus concurrent models in predicting total spending is summarized in appendix table A3 which will be of interest to those with experience using risk adjustment models.

## Section 3: Data

We analyze data on plan enrollment and insurance claims from the Truven Health Analytics MarketScan® Research Databases 2007-2011. Our base eligibility sample contains 5.1 million eligible people ages 21 to 64 who are continuously enrolled for five full years, resulting in 25 million enrollee years. The decision to require five years of continuous enrollment reflects our desire to control for enrollee taste variation, which we do with fixed effects for each person. The focus on adults is appropriate so that we have a more homogeneous set of treatment patterns. Each enrolled person is linked to his/her insurance claims incurred over the period, altogether encompassing 720 million inpatient and outpatient claims and 1.6 million inpatient admission records. We exclude pharmacy claims in the analysis because drug-level decisions differ from inpatient and outpatient visits and likely require a distinct model. Diagnoses and spending are assigned to spell months based on the from date of service for each procedure, not using thru (or ending) dates of service (as is done with Medicare Advantage risk adjustment), and not assigning services to the beginning of an episode (as was done by Keeler et al (1988)) or the date of discharge (as is implicitly done by DRGs). Since most claims for procedures are for a single date, we feel that the service date best reflects the timing when a patient and doctor are likely to know of an illness or condition.

A significant challenge in using the MarketScan (MS) data is dealing with missing values on key variables, namely employer, health plan IDs, provider IDs, and provider county (used for price

deflation). Since risk adjustment models require only age, gender, diagnoses, type of service and provider specialty information, which are generally available on all claims, we were able to use the full sample of 5.1 million to estimate risk adjustment models at the treatment spell level. We also used the full sample of enrollees to calculate US average prices by procedure for pricing intensity. Further details on sample selection and variable definitions are provided in Appendix B. Basically, for missing employer information, we assigned consumers to the same employer as observed during a previous or subsequent year since the only reason they are in the database is because they are with a particular employer. For providers with a missing county (which we used only for price deflators) we use the provider's state rather than the provider's county. We estimate choice models of health plan type as offered by an employer in a given year rather than the choice specific health plans to circumvent missing health plan IDs.

Prices were inflated into 2011 dollars. Geographic adjustment was done at the county level using the Medicare program's Geographic Practice Cost Index (GPCI) to convert into national prices. We also calculated the US average price for each procedure and used these averages to calculate what payments would have been if US average prices had been used instead of the actual prices. The ratio of the real price to the US prices, weighted by quantities of services in a spell is our measure of pricing intensity.

Patients were assigned to PCPs using the algorithm of Song et al (2012), as also implemented in Vats, Ellis and Ash (2013). This algorithm assigns patients to actual providers seen using provider IDs, provider specialty, procedure codes and type of service categories. Since we use monthly rather than annual periods, there is less of a need for multiple tie-breaking rules.[3] The only provider specialties assigned as PCPs were internal medicine, family practice, pediatrics, geriatrics, obstetrics/gynecology, medical doctor (not elsewhere classified), and "multispecialty practice." Switching between different PCPs were rare, occurring in 4 percent of all spell months and 41 percent of treatment spells with positive spending with at least one visit to a PCP, suggesting that PCP assignment is reasonably persistent. Further details about PCP assignment are provided in Appendix B.

Since primary care during pregnancies will differ from care at other times, we excluded all spells in which there was a pregnancy diagnosis, both concurrently or during the prior twelve month base period in which diagnoses are recognized for each spell. We also excluded spells with negative spending, spells continuing beyond our sample period, and any spells involving capitation payments because total payments will be unreliable.

## Section 4: Econometric Issues

Before moving on to the empirical results, we first discuss four econometric issues. First, it is well understood that health plan choices are endogenous, and reflect expected health care

[3] It would be preferred to have PCPs as chosen by consumers or providers, however such information would rarely be available for plan types such as HDHP and CDHP emphasizing consumer choice.

spending (Manning et al, 1987). Estimates of response to insurance plans that do not control for this endogeneity will be seriously biased. We control for this endogeneity by using a two stage least squares estimation in which our instruments are derived from a preliminary logit model estimating the probability of selecting a given plan type in a given year. Since we do not have plan specific premiums, benefits, or coverage features, we capture these plan features with a plan specific dummy for each plan type offered by each employer in each year. Fortunately, our employers vary dramatically in the menus of plans offered, and this first stage model has considerable predictive power ($R^2$=.393), even without including prior year utilization or diagnostic information. We include only enrollee demographics in this logit estimation, together with knowledge about the plan type choice set offered by each employer in each year.

The second challenge is that employers do not randomly decide whether to offer new health plan types: firms with healthier enrollees are much more likely to innovate and offer EPOs, CDHPs and HDHPs. All of our models use employer*year*family coverage fixed effects to absorb the underlying employer characteristics that may make them more or less likely to offer innovative plan types. We identify the effects of plan innovations by the change in coverage for continuously eligible households. Individuals who do not change plan types in our sample are uninformative about the effects of plan type on decisions.

A third challenge is that even the best risk adjustment models cannot capture all of the variation in health illness burden, nor consumer tastes which will also influence consumer choice of treatment options determined jointly by consumers and providers. We use individual and PCP fixed effects to capture these unobserved values, and estimate changes in treatment decisions made by each individuals over time.

A fourth econometric challenge is that decision errors are correlated at the employer, year, person and PCP level, so that calculating standard errors that will not be misleading is important. Given our large sample sizes, augmented by modeling at the month level, it is tempting to interpret the enormous t ratios (which can be over 2000) as highly significant. But this would be wrong. The relevant unit of observation in our study is the health-plan-year, of which we have 1324 distinct offerings. Following Donald and Lang (2007) we use a two-step estimator that first uses our millions of spell months records to generate coefficients on a set of dummies reflecting the interaction of each combination of plan type*employer*year*family coverage (each combination is also known as a group). An inference is then made with secondary regressions that use these regression coefficients as the dependent variable and regress on variables that vary at the group level.

With the exception of the plan choice models, which are estimated using a logit model separately for each employer*year*family coverage type, we use a linear model for all of our dependent variables $Y_{ipjt}$ that can be written as follows.

$$Y_{ipjt} = f(RCC_{it,} RCC_{it-1}) + \mu_i + \lambda_p + \psi_j + \delta_t + \varepsilon_{ipjt} \qquad (2)$$

where $Y_{ipjt}$ denotes each of our dependent variables for individual i with health plan p and PCP j in spell month t. $\mu_i$, $\lambda_p$, $\psi_j$, , and $\delta_t$ are individual, health plan, PCP, and spell month (time) fixed effects respectively, and $\varepsilon_{ijpt}$ is an error capturing other unobserved or random terms.[4]

To accommodate the large number of fixed effects while using two stage least squares, we use the following estimation algorithm. For each study sample, we first difference out the enrollee (N >1 million) (and as needed the PCP (N>100,000)) fixed effects from various dependent and independent variables including our instruments for endogenous plan type choice (namely the fitted probabilities of choosing each plan type). In the second step, we estimate two stage least squares models using as dependent variables the set of coefficients on the dummy variables for each plan type*employer*year*family coverage combination. Since these coefficients are only identified when there is a change in a coverage choice by an employer and a change in choices by one or more consumers, our final estimation sample is only on the order of 1324 observations. Since the first step estimated coefficients are less precise for plans with small enrollments, we use feasible generalized least squares (FGLS) to calculate our coefficients with corrected standard errors. Specifically, in our second-step estimation, we regress the squared residuals from the first stage of a TSLS model on a constant and 1/N and then use the inverse of these fitted values to weigh observations in the second stage, based on the notion that less weight is given to observations with smaller enrollments.

## Section 5: Results

### Section 5.1 Summary statistics

Table 1 provides summary statistics by plan type at the employee-year level for seven plan types included in this study. (Further descriptive statistics for the full and restricted sample are provided in Appendix Table A-2.) This table shows the rich set of plan types offered by the 66 employers in our full sample, with seven employers offering EPOs. HMOs and PPOs remain the two most common plan types, together accounting for more than 60 percent of the employee years in our sample. Employee age, family size, and the sum of relative risk scores among adults in the household all signal significant risk differences across plan types. Comprehensive plans have an average age that is eight years older than EPOs, which have the youngest adult enrollees. Although EPOs, HMOs, CDHP and HDHP plans have above average family sizes, they have below average risk scores, suggesting that individuals in these plans are healthier than average given their age.

As shown in Figure 2, there have been sharp changes in enrollments in different health plan types even over the four years being studied here. PPOs grew by five percentage points from

[4] We also explored including county fixed effects, but this had essentially no effect on the results. County FE are only identified by people who changed counties, which is rare in our sample (1.7% per year) and county variables are missing altogether from the 2011 sample. Therefore we dropped them from our primary analysis. Results from adding county fixed effects are presented in the appendix.

2008 to 2009 before declining by the almost same amount in 2011. CDHPs increased their market share by eight percent, while HDHPs saw a three percentage point increase. Declines were experienced by HMOs and their closely related POS plans. Two plan types, EPOs and Comprehensive plans (COMP) held their enrollments at a nearly constant one percent market share throughout the period.

### Section 5.2: Choice of health plan

We first model enrollee choice among the health plan types offered by their employer. We estimate two types of choice models, models that include risk scores that capture potentially endogenous risk scores, and reduced form models that omit risk scores and only use enrollee demographic information to predict plan choices. Fitted plan type choices from the reduced form are used as instruments in our two stage least squares (TSLS) models of subsequent treatment decisions. Our basic specification is

$D_p$ ($Z_{it}$) = 1 if the consumer i chooses plan type p

where $Z_{it}$={employer, employee age, employee gender, family size, family/single coverage, spouse, new baby flag, and RRS t-1}.

Given the large sample size and lack of plan-specific features such as premiums, we estimate this process using a multinomial logit with a separate plan specific dummy for each employer-year-plan type. Employees with only one insured enrollee were modeled as single plans, and those with multiple enrollees were modeled as family plans. Altogether, 242 separate logit models were estimated at the employer*year level, separately for single and family plan type choices. Since it is unclear what risk score best predicts health plan choices, we experimented with eight different relative risk scores that predict spending using prior year diagnoses to predict health plan choice. (See Appendix Table A-4.) Empirically, the single best risk score predicting plan choice is the prospective, all diagnoses RRS, summed up for all adults in the household. Without exploring alternative topcoding thresholds, we topcoded this risk score at the 95th percentile (RRS=4.824) to reduce the influence of extremely high risk score individuals in predicting health plan choice, hypothesizing that risk scores differences above this level (predicting household spending over $20,000 per year) do not differentiate plan choices well. The topcoded risk scores were consistently superior to the un-topcoded scores in predicting plan choices.

We conducted a series of likelihood ratio tests on the full sample, (Appendix Table A-4) and all variables were found to be statistically significant at high levels. It may be relevant to future researchers to note that in terms of predictive power by far the most predictive information of health plan choice is whether an employer offers that health plan type.[5] Conditioning on being

[5] This may seem obvious, but most survey databases do not contain the full set of health plan alternatives offered, and hence this information is generally missing.

offered, the employee's age, the risk score, sex, family size, spouse, and whether there is a baby added in the past year are significant in decreasing order of predictive power.

A separate paper could potentially be written about choice inertia, and patient reluctance to change plans using our framework. However without premiums, benefit features and information about the panel of providers included in each network, this would be difficult at best. We will note that irrationality and imperfect optimization implies an asymmetry between those switching plans and those not switching: Switchers into the new health plan will likely have unrepresentative positive attitudes about these new plan types, while those who do not switch are more likely to have negative sentiments about new plan types (such as by knowing that they have an undetected chronic condition or are risk takers).

Although our structural specification predicting plan type choice includes lagged relative risk scores, our testable hypothesis is that (lagged) diagnostic coding may be endogenous to the health plan choice. To avoid using instruments that are correlated with expected spending, we use as instruments only the reduced form, fitted probabilities of a patient choosing a plan type without using lagged risk scores or diagnoses in subsequent models of treatment choice and levels of spending. Employers who offer only one plan type are given an assigned probability of one to every employee, included with the logit estimated probabilities. The actual plan type choices and our instruments are reasonably correlated, with an R-square of .393, which is usually excellent for an IV (i.e. relevance).

### Section 5.3: The decision to seek treatment

Each month, the first choice is whether to begin or continue a spell of treatment. We call this treatment choice "making a visit", even though treatment may also involve visits to laboratories, specialists, hospitals or emergency departments. Separate models for different types of visits conditional on a visit (as in Manning et al, 1987) can be readily estimated, but we initially pool all types of services and all spell types, $S_{it}$, into one model, and capture choice of setting as part of the treatment intensity decision that we model below.

$D_t (X_{it-1}, S_{it-1}) = 1$ if the consumer i seeks any care in the spell month t

In addition to the $S_{it-1,}$ type of spell in the previous period, we include among the explanatory variables, $X_{it-1,}$ including age and sex group dummies, monthly time dummies, employer*year*family coverage fixed effects, enrollee, and a prospective model risk score predicting total spending using the prior twelve months of diagnoses. Since we cannot observe concurrent diagnoses without knowing that a visit occurred, we use only predetermined information to predict visits. The mean probability of a visit in a monthly spell is only .32, so estimated effects of health plans on these decisions are in fractions of a visit per month.

Table 2 shows the detailed regression results for the unconditional visits using both OLS and TSLS. As further controls are added in OLS, most of the coefficients become smaller, suggesting that much of the apparent reduction in visits from alternative plan types is related to selection

rather than true savings. Model (3), which achieves statistical significance of all of the six plan types as differing from PPOs, with the largest effects captured by HDHPs-. Specifically, relative to PPOs, HDHPs have an effect of lowering the probability of unconditional visits by 7.7 percentage points. In fact, the OLS results show that compared to PPOs, all other plans tend to discourage patients from making a visits, regardless of whether these other plans feature a relatively higher cost sharing (like CDHPs and HDHPs or more restrictive provider network (like EPOs and HMOs) or even higher generosity (like Comprehensive plans). Also shown in the final three columns of Table 2 are results from further controlling for the endogeneity of health plan choice, Almost all of the health plan effects on the probability of visits become insignificant. Moreover, coefficients become unstable as additional covariates are added. We speculate that this is because most of the explained variation in plan choice is related to the employer offering, with only modest predictive power of independent variables such as family size, spouse, employee age, and a flag for a baby. Once employer or person fixed effects are included, this undermines the strength of our IV, giving us weak power.

Having illustrated the health plan effects on any type of visits, we now look at the three types of visits separately. As previously discussed, we used N, V and C to denote no visit, new visit, and continuation spells and we are interested in the probabilities of a consumer initiating or continuing a visit conditional on whether or not he/she pays a visit in previous spell month, denoted as $Pr(V|N_{-1})$, $Pr(C|V_{-1})$, and $Pr(C|C_{-1})$, where the negative sign denotes the treatment spell of the consumer in the prior month. Results from the OLS models with employer*family fixed effects and risk adjustment but excluding individual fixed effects (corresponding to model (3) in table 2) are shown in the appendix table A-6, and summarized graphically in Figure 3. Effects that are statistically significant are marked with asterisks according to their level of significance. Our data show that relative to the omitted group of PPOs, overall HDHPs show the greatest reduction in the probability of a visit, followed by HMO and CDHP plans, while POS plans show the smallest reduction in visits overall. Once again, plans in generally show a decreasing numbers of visits relative to PPOs whenever the effect is significant. Whereas HDHPs have the largest overall impact on all types of visits, HMOs are relatively more effective in reducing continuation visits and Comprehensive plans display their main impact on cutting new provider contacts.

### Section 5.4: Decomposition of total payment variation

A major objective of this study has been to understand the factors that contribute to the enormous variation in spending conditional on a patient deciding to seek treatment for a spell of illness. We decompose the log of total payment on procedures in monthly treatment spells at the individual level (conditional on payment being positive) using the following algorithm. Decomposing explained variation is not feasible to do in one regression because of the enormous number of individual, employer, and PCP fixed effects involved. Therefore instead we calculate the incremental contribution of each class of variables when added one at a time (which by SAS and

elsewhere is often called the Type III explained variation). Results are shown in Figure 4 for one sequence of adding sets of variables.

At the onset, before any other modeling, we deflated payments to remove time trend and geographic variation, which together explain one half of one percent of variation. Our proxy for health plans, which are health plan types interacted with employer, year and family/single coverage dummies, only explains an additional 1.1 percent more of the payment variation. Individual fixed effects explain an additional 20.8 percent of total variation, and pick up not only taste, but also chronic conditions that are persistent across our four year sample period. Remarkably to us, adding PCP fixed effects only explains an additional 1.6 percent of the payment variation. Pricing intensity variation, which captures deviations from the national average, explains an additional 4.7 percent of variation. Our measure of diagnostic severity, which uses a coarse set of condition categories from both the prior 12 months and the concurrent 12 months, explains an additional 17.4 percent of the variation, even after including patient fixed effects. Finally, in addition to the severity measure, our coding intensity measure, which captures the variation in detailed coding assigned to patients by providers beyond information already captured by our illness severity, explains an additional 2.6 percent of total variation in log of monthly spending. Altogether, the seven factors just described explain 48.7 percent of total variation in the log of spending on procedures in monthly spells of treatment.

### Section 5.5: Association of plan type with payment variation

We now examine how health plan types are associated with levels of health care spending conditional on a patient deciding to make a visit. Recall that we are only modeling spending on procedures; facility charges - which are much harder to price or to predict– are excluded from our primary analysis here. An important rationale for this focus is that our calculation of national average prices is more precise for procedures than for facility charges. Analysis is conducted by estimating OLS and TSLS models of the log of spending and logs of three intensity measures at the level of monthly spells conditional on spending being positive.

Because there are too many individual and PCP fixed effects to include in a regression model, we absorb them by differencing them out in all of our models. As discussed in the econometrics section, we estimate models predicting total spending on procedures and our three intensity measures using a two step procedure, with TSLS done in the second step to control for endogenous plan choice. Corresponding to Table 3, seven specifications are presented, for varying sets of covariates. The first column uses simple OLS, with only monthly time dummies. Adding employer fixed effects and risk adjustment has only a modest effect on total payments for procedures. Once patient fixed effects are added in column 4, health plan treatment effects become larger in absolute magnitude, reaching .272 , not smaller, hinting at the possibility that reductions in visits  while subsequent models use TSLS with successively more fixed effects and control variables. Here again, we see that OLS suggests much larger effects of most plan types relative to PPOs, and this pattern persists when only employer*year*family and risk adjustment scores are included. Plan type effects become much smaller in almost every case once individual

fixed effects are included. Not shown here is the effect of adding in a further PCP fixed effect. We chose not to include PCP fixed effects in our final table because there is reason to believe that the PCP fixed effects absorb some of the impact of plan incentives: one impact of changing plans and responding to cost sharing or restricted provider choice could be to switch to a lower cost provider. Overall, differences when adding in PCP fixed effects are relatively small.

Our preferred specifications appear in column four, which use employer*year*family, a customized risk score intended to capture consumer relevant information, and individual but not PCP fixed effects. The log of real procedure payments model as well as each of the three intensity measures has relatively small coefficients. As before, the TSLS results find almost all coefficients to be statistically insignificant, which we attribute to the issues discussed in the previous section.

Structured in the same manner as the figure for any visit, Figure 5 help us visualize our regression results. We chose to focus on column 4, the OLS model with risk adjustment and patient and employer fixed effects. Model 4 results find relatively increase in costs for CDHP and COMP relative to PPOs, explained by higher prices and greater treatment intensity. Our measure of coding intensity finds statistically significant differences for HMOs and POS plans. First, relative to reductions in numbers of visits, particularly new visits, choosing a lower price provider is a relatively unimportant response to health plan level incentives. The only two health plans that show significant effects are high deductible plans emphasizing consumer choice (esp. HDHPs) and comprehensive plans with little cost sharing. Relative to PPO plans, these two plans are associated with higher prices after controlling for employer, individual and time fixed effects.

## Section 6 Discussion and Conclusions

This paper examines variation in utilization and payments for a monthly spell of treatment, which is arguably a more appropriate unit for studying decision making than annual or monthly periods. In part as a proof of concept, we have taken a decidedly “macro view” by examining all procedures by all types of providers for all types of diagnoses rather than using only a narrow subset of services, providers and conditions. Our only substantive exclusions were to focus on adults, omit pregnancies, and model only spending on procedures rather than total payments. Spells of treatment, as a readily available alternative, are well suited to these more micro focused studies, but we postpone those more focused analyses for a future paper.

There is enormous interest in figuring out why health care costs vary so much, how good performance can be detected and rewarded, and whether various demand or supply side innovations can help promote lower costs and better performance. Careful risk adjustment, together with controls for pricing variation suggest that nearly half of the spell month variation in payments (when payments are positive) can be explained, reducing the amount of noise that makes answering these questions so difficult. Another advantage of shorter periods of modeling

is that potentially feedback can be provided to payers, policymakers, patients and providers in a more timely manner, although this will also require greater data sharing and standardization.

The classic summary of the RAND health insurance experiment was that "insurance primarily affects individual decisions to seek treatment (episode frequency), but has only minimal effects on episode costs" (Buchanan, et al, 1991). Our analysis finds that health plan types have other significant effects besides changing visit rates, reducing payments in HDHPs in most specifications and HMOs overall in those with the most controls.

Our results suggest that estimates of cost savings of major health plan innovations, such as EPOs, CDHP and HDHP, are remarkably difficult to pin down, even with careful risk adjustment and one of the largest samples of health plans employers, and consumers that we have seen. Results seem particularly sensitive to the inclusion of individual fixed effects. This reinforces that health plans differ in who they attract as much as how they change behavior among those who enrolls. This is not a new finding, but the magnitudes of it seem large.

This paper has only assigned spells of treatment to primary care providers (PCPs), although in principle assignment to a broader array of providers will be feasible and attractive. We also have continued to use only claims based diagnostic information for risk adjustment, although pharmacy information, and ideally consume socioeconomic variable that affect access and information are also important. We use here "off the shelf" risk adjustment models but look forward to being able to broaden the information available, and customize the models to predict diverse outcomes.

## References

Altman, Daniel, David Cutler, Richard Zeckhauser (2003) "Enrollee mix, treatment intensity, and cost in competing indemnity and HMO plans," Journal of Health Economics, 22(1): 23-45.

Aron-Dine, Aviva, Einav, Liran, Finkelstein, Amy, and Cullen, Mark R. (2012). "Moral Hazard in Health Insurance: How Important Is Forward Looking Behavior?" NBER Working Paper No. 17802.

Ash, Arlene S., and Randall P. Ellis (2012). "Risk-Adjusted Payment and Performance Assessment for Primary Care," Medical Care, 50(8): 643-653.

Ash, Arlene S., Ellis R.P., Gregory C. Pope, John Z. Ayanian, David W. Bates, Helen Burstin, L.I. Iezzoni, E. McKay, W. Yu, (2000). "Using Diagnoses to Describe Populations and Predict Costs," Health Care Financing Review, Spring 21(3): 7-28.

Berenson RA, Rich EC. (2010). "US approaches to physician payment: the deconstruction of primary care," J Gen Intern Med, 25(6): 613-618.

Bitton A, Martin C, Landon BE 2010. "A Nationwide Survey of Patient Centered Medical Home Demonstration Projects," Journal of General Internal Medicine, 25(6): 584–592. doi:10.1007/s11606-010-1262-8. PMC 2869409. PMID 20467907.

Buchanan, J.L. Keeler E.B., Rolph J.E. and Holmer M.R. 1991. "Simulating Health Expenditures under Alternative Insurance Plans," Management Science. 37(9): 1067-1090.

Cutler, David M., Finkelstein, Amy, and McGarry, Kathleen, (2008). "Preference Heterogeneity and Insurance Markets: Explaining a Puzzle of Insurance," American Economic Review, Papers and proceedings, 98(2): 157-162.

Dafny L., S, Ho K. and Varela M. 2013. "Let them Have Choice: Gains from Shifting away from Employer-Sponsored Health Insurance and Toward an Individual Exchange," American Economic Journal: Economic Policy. 5(1): 32-34.

Donald, S. G., and K. Lang. 2007. "Inference with difference-in-differences and other panel data," Review of Economics and Statistics 89(2): 221–233.

Einav, Liran, Amy Finkelstein, Stephen P. Ryan, Paul Schrimpf, and Mark R. Cullen. 2013. "Selection on Moral Hazard in Health Insurance," American Economic Review, 103(1): 178-219.

Ellis, R.P., and A.S. Ash. 1989. "The Continuous Update Diagnostic Cost Group Model," Final report prepared for the Health Care Financing Administration, June.

Ellis, Randall P. (1986). "Rational Behavior in the Presence of Coverage Ceilings and Deductibles," The Rand Journal of Economics, Summer, 17(2): 158-175.

Ellis, Randall P. (2012) "Five questions for health economists," International Journal of Health Care Finance and Economics 12(3): 217-233.

Ellis, Randall P. (2008). "Risk adjustment and predictive modeling: Concepts and applications," in Lu, Mingshan and Egon Jonsson. (eds.) Financing Health Care: New Ideas for a Changing Society. Wiley-VCH publishers Weinheim, Germany. 177-222.

Ellis, Randall P., Denzil G. Fiebig, Meliyanni Johar, Glenn Jones, Elizabeth Savage (2013) "Explaining Health Care Expenditure Variation: Large Sample Evidence Using Linked Survey and Health Administrative Data," Health Economics, 22(9): 1093-1110.

Ellis, Randall P. and Thomas G. McGuire (2007). "Predictability and Predictiveness in Health Care Spending," Journal of Health Economics, 26(1): 25–48.

Finkelstein, Amy, (2014). "Moral Hazard in Health Insurance: Developments since Arrow," Columbia University Press.

Geruso, Michael and Timothy J. Layton (2014). "Risk Selection, Risk Adjustment, and Manipulable Medical Coding: Evidence from Medicare," Unpublished.

Hellinger, Fred J. (1994). "Selection bias in HMOs and PPOs: a review of the evidence," Inquiry: a journal of medical care organization, provision and financing, 32(2): 135-142.

Keeler, E.B., W.G. Manning and R.B. Wells (1988). "The demand for episodes of mental health services," Journal of Health Economics, 7(4): 369-392.

Keeler, E.B., and Rolph. J.E. (1988) "The demand for episodes of treatment in the Health Insurance Experiment," Journal of Health Economics, 7(4): 337-367.

Manning, Willard G, et al. (1987). "Health insurance and the demand for medical care: evidence from a randomized experiment," American Economic Review, 77(3): 251-277.

Pope, Gregory C, John Kautter, Randall P. Ellis, Arlene S. Ash, John Z. Ayanian, Lisa I. Iezzoni, Melvin J. Ingber, Jesse M. Levy, and John Robst. (2004). "Risk Adjustment of Medicare Capitation Payments Using the CMS-HCC model," Health Care Financing Review, Summer 25(4): 119-141.

Reid, Robert J. and Trescott, Claire (2013). "Medical Home Features Small Panels, Long Visits, Outreach, and Caregiver Collaboration, Leading to Less Staff Burnout, Better Access and Quality, and Lower Utilization," Agency for Healthcare Research and Quality, 2013-05-08.

Robertson C, et al. (2012). "Effect of financial relationships on medical professionals- a review of the evidence," Journal of Law, Medicine and Ethics, 40: 452–466.

Robinson, James C., Timothy T. Brown (2013). "Increases In Consumer Cost Sharing Redirect Patient Volumes And Reduce Hospital Prices For Orthopedic Surgery," Health Affairs, 32(8):1392-1397.

Song, Zurui and Bruce E. Langdon (2012). "Controlling Health Care Spending — The Massachusetts Experiment," New England Journal of Medicine, 366: 1560-1561.

Song, Zurui, Dana G Safran, Bruce E Landon, et al. (2011). "Health Care Spending and Quality in Year 1 of the Alternative Quality Contract," New England Journal of Medicine,365: 909-918.

Van de Ven, Wynand .P.M.M. and Randall P. Ellis (2000). "Risk Adjustment in Competitive Health Plan Markets," Chapter 14 in A.J. Culyer and J.P. Newhouse, (ed.) Handbook in Health Economics, North Holland. 755-845.

Vats, Sonal, Arlene S. Ash and Randall P. Ellis. (2013). "Bending the Cost Curve? Results from a Comprehensive Primary Care Payment Pilot," Medical Care, 51(11): 964-969.

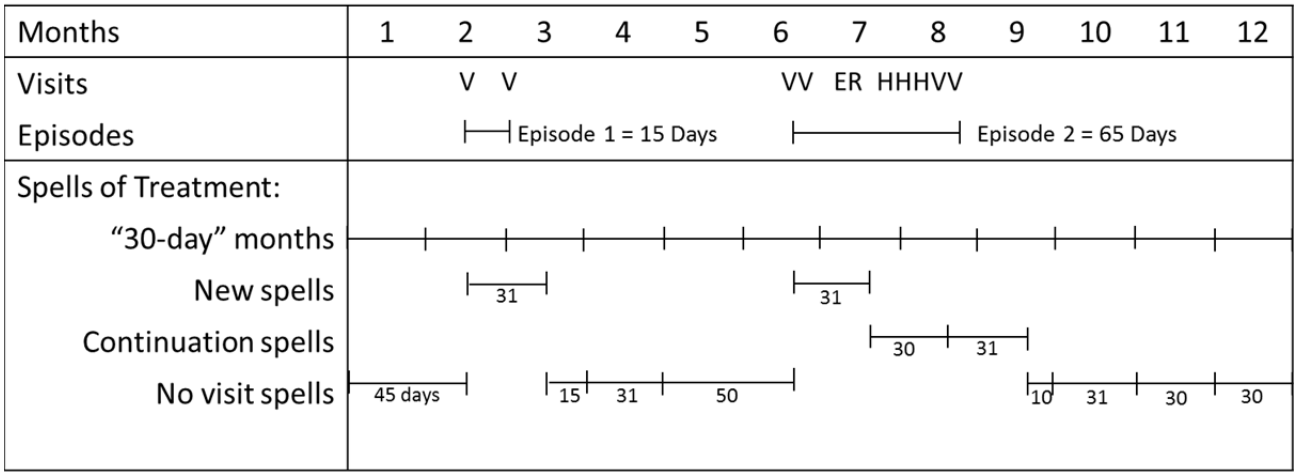
Verisk Health, Inc. 2013. DxCG Intelligence User Guide v4.2.1, Waltham MA.

Wennberg, John E, Staiger, Douglas O., Sharp, Sandra M et al. (2013). "Observational intensity bias associated with illness adjustment: cross sectional analysis of insurance claims," British Medical Journal, 346 doi: http://dx.doi.org/10.1136/bmj.f549.

Wennberg, David E., Sandra M. Sharp, Gwyn Bevan, Jonathan S. Skinner, Daniel J. Gottlieb, and John E. Wennberg. (2014) "A population health approach to reducing observational intensity bias in health risk adjustment: cross sectional analysis of insurance claims," BMJ: British Medical Journal 348.

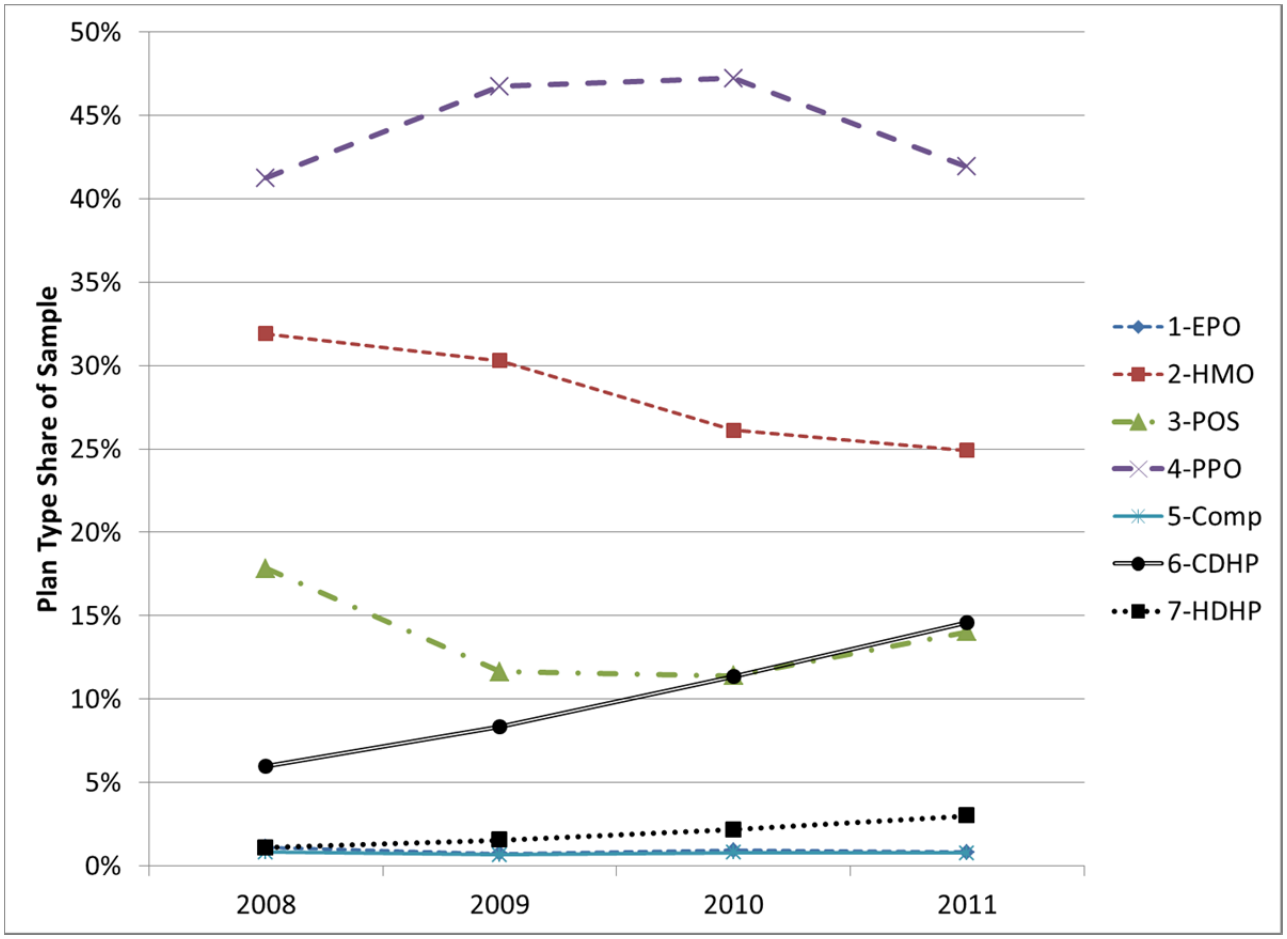
Wennberg, John, and Alan Gittelsohn. (1973). "Small Area Variations in Health Care Delivery A population-based health information system can guide planning and regulatory decision-making," Science 182, no. 4117: 1102-1108.

**Figure 1: Example of Treatment Spell Assignment**



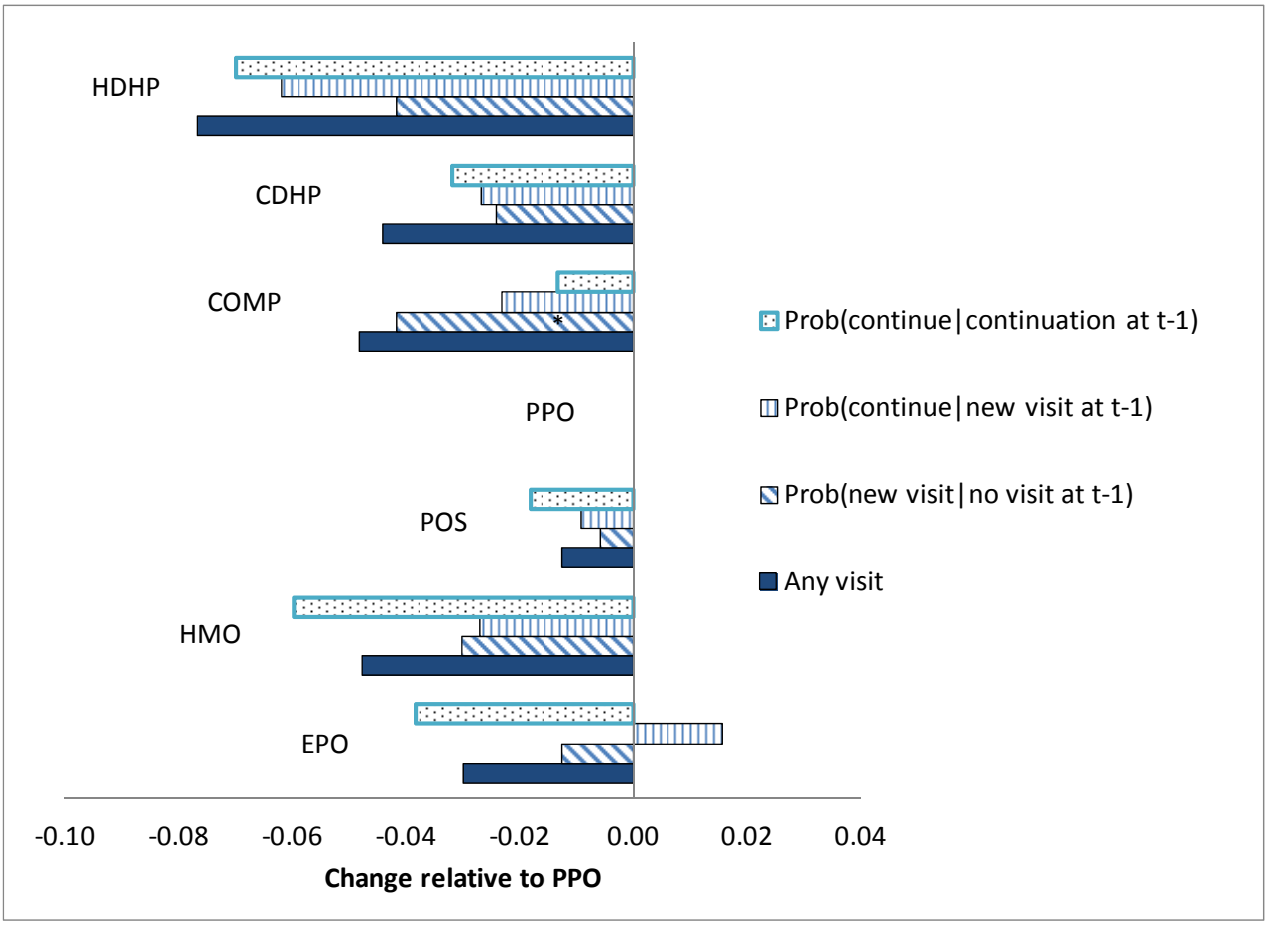

Notes: Spell assignment begins by dividing up the year into 12 months of 30 or 31 days. New spells begin on the day of a first visit after a gap of "30 days" since a previous provider contact. Continuation spells follow new or other continuation spells. No visit spells fill in around new and continuation spells and can have variable lengths of from 1 to 60 days.

Figure 2 Market shares of seven health plan types among the employer subset of enrollees 2008-2011



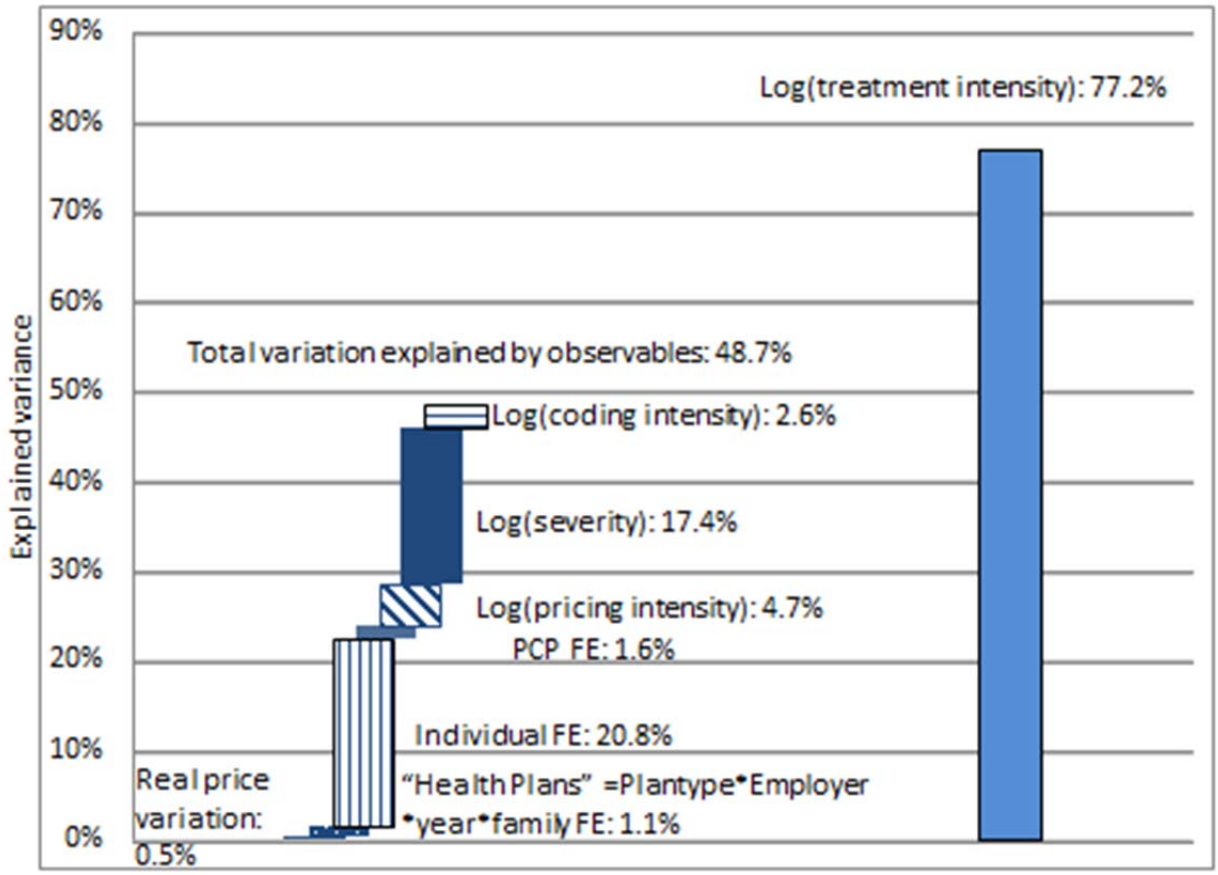

Notes: Market shares are calculated at the patient level. Acronyms of health plan types are explained in text and appendix table A-1.

Figure 3 Association between plan type and the probability of receiving treatment during a spell month.



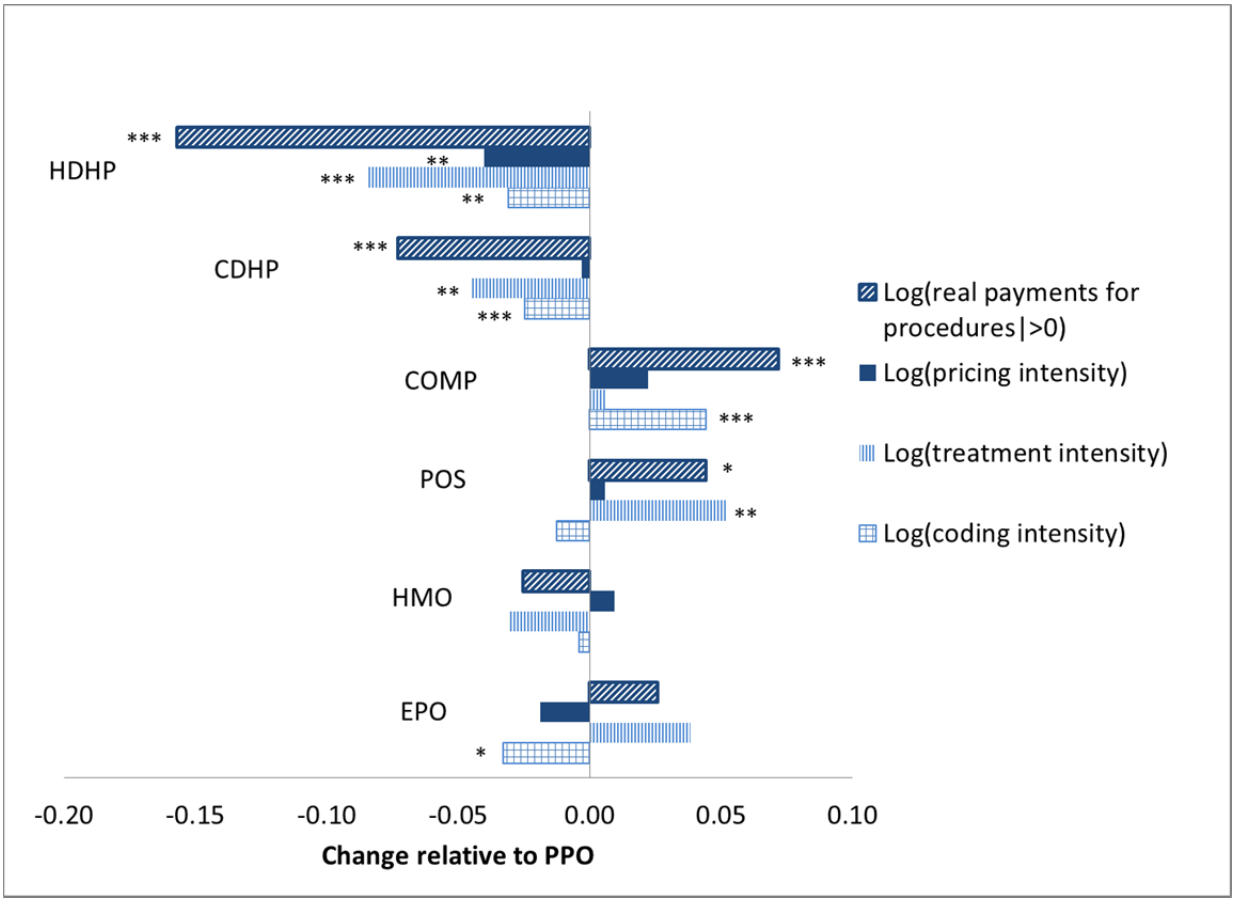

Notes: Results shown are from four OLS regressions on samples at the treatment spell level. All the models include age sex groups, employer*year*family coverage, spell month time dummies, and HCC based RRS from disease recorded during the previous 12 months. Full regression results are in Table 2.

Figure 4 Explained variance in log of spell-month payments for procedures given that they are positive




Note: Figure shows the proportion of variation in the log of monthly spending on procedures as calculated by Type III sums of explained variation, which is not invariant to the order in which variables are added. Variables were added successively from the bottom to top to explain total variation, and the incremental variance explained is shown. Results use the sample of spells of treatment that can be assigned to employers and PCPs, N=20 million.

Figure 5. Association between plan type and real payment (when >0) and intensity measures




Notes: Results shown summarize coefficient from OLS regressions at the treatment spell level with the dependent variables shown in the legend. Real payments denotes payments for all procedures. Sample only uses months in which spending is positive (N=20,090,330). Since PPOs are the omitted plan type, all results reflect differences from PPOs. All models include age-sex groups, employer*year*family coverage dummies, spell index dummies, and HCC based RRS from disease recorded during the previous 12 months. Full regression results are in column (3) of Table 3.

Table 1 Summary statistics by plan types in employer sample

| Plan Type | Acronym | Number of employers in sample | Number of employer-years in sample | Number of employee-years | Employee age | Family size | Sum of adult prospective Relative Risk Scores |
|---|---|---|---|---|---|---|---|
| Exclusive Provider Organization | EPO | 7 | 15 | 40,694 | 44.6 | 2.6 | 1.675 |
| Health Maintenance Organization | HMO | 38 | 147 | 1,924,987 | 45.6 | 2.5 | 1.804 |
| Point of Service | POS | 32 | 101 | 735,592 | 45.1 | 2.0 | 1.505 |
| Preferred Provider Organizations | PPO | 58 | 239 | 3,547,872 | 46.0 | 2.4 | 1.804 |
| Comprehensive | COMP | 36 | 124 | 184,716 | 52.7 | 2.0 | 1.873 |
| Consumer Directed Health Plan | CDHP | 26 | 99 | 525,505 | 45.4 | 2.4 | 1.723 |
| High Deductible Health Plan | HDHP | 14 | 46 | 107,593 | 46.2 | 2.5 | 1.608 |
| ALL | | 66 | 771 | 7,066,959 | 45.9 | 2.4 | 1.765 |
| | | | | | (10.12) | (1.41) | (1.26) |

Notes: Results shown are for employer subsample of enrollees, with means calculated for the employee, not patient. Mean family size include children who are not otherwise included in the anlaysis of adults here. Family size is top coded at 10. Relative risk scores are from the DxCG software, summed up for any adults in the family, without any rescaling or recalibration to this data. See Appendix table A-4 for further discussion. Excluded are point-of-service plans with full or partial capitation, and plans with very small enrollments (N<50). Number of employer-years is the sum of how many employers offer at least one plan of this type in a given year. Multiple plans of the same plan type are counted only once. Standard deviations are in parentheses.

Table 2 Regression results showing health plan effects on unconditional visits

| | (1) OLS | (2) OLS | (3) OLS | (4) OLS | (5) TSLS | (6) TSLS | (7) TSLS |
|---|---|---|---|---|---|---|---|
| Employer*family FE | | X | X | X | X | X | X |
| Risk adjustment | | | X | X | | X | X |
| Indivdiual FE | | | | X | | | X |
| Visits (all types) | | | | | | | |
| (N=62,971,626) | Coef | Coef | Coef. | Coef. | Coef. | Coef. | Coef. |
| | (Std. Err.) | (Std. Err.) | (Std. Err.) | (Std. Err.) | (Std. Err.) | (Std. Err.) | (Std. Err.) |
| EPO | -0.026 | -0.029 * | -0.030 ** | -0.022 * | 0.002 | -0.005 | 0.002 |
| | (0.02) | (0.02) | (0.01) | (0.01) | (0.07) | (0.05) | (0.02) |
| HMO | -0.066 *** | -0.077 *** | -0.048 *** | -0.011 | 0.015 | 0.010 | 0.026 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.09) | (0.06) | (0.02) |
| POS | -0.008 | -0.012 | -0.013 ** | 0.001 | 0.066 | 0.042 | 0.021 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.06) | (0.04) | (0.02) |
| COMP | 0.008 | 0.005 | -0.048 *** | -0.035 *** | 0.189 *** | 0.048 | 0.044 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.06) | (0.05) | (0.04) |
| CDHP | -0.082 *** | -0.083 *** | -0.044 *** | -0.023 *** | -0.057 | -0.034 | -0.049 *** |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.06) | (0.04) | (0.02) |
| HDHP | -0.119 *** | -0.126 *** | -0.077 *** | -0.033 *** | 0.039 | 0.034 | 0.100 ** |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.09) | (0.06) | (0.05) |

Notes: Results from 7 regressions using grouped data are presented. Spells were first summarized to the plan type*employer*year*family/single coverage and then used for grouped regressions.

Table 3 Detailed regression results showing health plan effects on real payments, and intensity measures

| | (1) OLS | (2) OLS | (3) OLS | (4) OLS | (5) TSLS | (6) TSLS | (7) TSLS |
|---|---|---|---|---|---|---|---|
| Employer*family FE | | X | X | X | X | X | X |
| Risk adjustment | | | X | X | | X | X |
| Indivdiual FE | | | | X | | | X |
| Log(real payments for procedures\| > 0) | Coef (Std. Err.) | Coef (Std. Err.) | Coef. (Std. Err.) | Coef. (Std. Err.) | Coef. (Std. Err.) | Coef. (Std. Err.) | Coef. (Std. Err.) |
| EPO | -0.008 | -0.005 | 0.026 | -0.003 | -0.139 | 0.010 | -0.223 |
| | (0.056) | (0.048) | (0.052) | (0.122) | (0.260) | (0.317) | (0.433) |
| HMO | -0.074 *** | -0.029 | -0.026 | -0.069 | -0.363 | -0.384 | 0.106 |
| | (0.024) | (0.020) | (0.022) | (0.063) | (0.358) | (0.449) | (0.358) |
| POS | 0.040 | 0.035 | 0.044 * | 0.107 | 0.016 | 0.024 | 0.108 |
| | (0.026) | (0.023) | (0.025) | (0.071) | (0.195) | (0.242) | (0.322) |
| COMP | 0.081 *** | 0.078 *** | 0.072 *** | 0.233 *** | 0.018 | -0.223 | 0.390 |
| | (0.026) | (0.022) | (0.024) | (0.073) | (0.335) | (0.446) | (0.568) |
| CDHP | -0.123 *** | -0.062 *** | -0.073 *** | 0.272 *** | -0.117 | -0.350 | -0.060 |
| | (0.026) | (0.023) | (0.025) | (0.083) | (0.182) | (0.213) | (0.269) |
| HDHP | -0.162 *** | -0.170 *** | -0.157 *** | -0.120 | -0.427 | -0.443 | 1.365 |
| | (0.035) | (0.030) | (0.033) | (0.091) | (0.500) | (0.709) | (1.257) |
| Log(pricing intensity) | Coef (Std. Err.) | Coef (Std. Err.) | Coef. (Std. Err.) | Coef. (Std. Err.) | Coef. (Std. Err.) | Coef. (Std. Err.) | Coef. (Std. Err.) |
| EPO | -0.042 | -0.018 | -0.019 | 0.026 | -0.034 | 0.010 | -0.143 |
| | (0.030) | (0.027) | (0.030) | (0.047) | (0.210) | (0.225) | (0.183) |
| HMO | 0.019 | 0.015 | 0.009 | -0.002 | -0.283 | -0.270 | 0.086 |
| | (0.013) | (0.011) | (0.012) | (0.024) | (0.296) | (0.319) | (0.140) |
| POS | 0.018 | -0.003 | 0.006 | 0.039 | 0.002 | -0.062 | -0.035 |
| | (0.014) | (0.013) | (0.014) | (0.027) | (0.159) | (0.173) | (0.133) |
| COMP | -0.015 | 0.025 ** | 0.022 | 0.102 *** | -0.184 | -0.291 | 0.167 |
| | (0.014) | (0.012) | (0.014) | (0.028) | (0.287) | (0.320) | (0.226) |
| CDHP | -0.028 * | 0.009 | -0.003 | 0.102 *** | 0.184 | 0.036 | 0.129 |
| | (0.014) | (0.013) | (0.014) | (0.032) | (0.143) | (0.151) | (0.109) |
| HDHP | -0.047 ** | -0.034 ** | -0.040 ** | -0.015 | -0.285 | -0.404 | 0.639 |
| | (0.019) | (0.017) | (0.019) | (0.035) | (0.446) | (0.511) | (0.543) |

Table 3 Detailed regression results showing health plan effects on real payments, and intensity measures (continued)

| Log(treatment intensity) | Coef | Coef | Coef. | Coef. | Coef. | Coef. | Coef. |
|---|---|---|---|---|---|---|---|
| | (Std. Err.) | (Std. Err.) | (Std. Err.) | (Std. Err.) | (Std. Err.) | (Std. Err.) | (Std. Err.) |
| EPO | 0.047 | 0.006 | 0.038 | -0.031 | -0.003 | 0.081 | -0.120 |
| | (0.045) | (0.042) | (0.046) | (0.093) | (0.182) | (0.254) | (0.302) |
| HMO | -0.068 *** | -0.038 ** | -0.031 | -0.012 | 0.064 | -0.139 | 0.152 |
| | (0.019) | (0.017) | (0.019) | (0.048) | (0.246) | (0.359) | (0.263) |
| POS | 0.038 * | 0.054 *** | 0.052 ** | 0.106 ** | 0.122 | 0.211 | 0.119 |
| | (0.022) | (0.020) | (0.022) | (0.054) | (0.137) | (0.193) | (0.231) |
| COMP | 0.023 | 0.012 | 0.006 | 0.118 ** | 0.165 | 0.003 | 0.394 |
| | (0.021) | (0.019) | (0.021) | (0.056) | (0.226) | (0.348) | (0.417) |
| CDHP | -0.049 ** | -0.050 ** | -0.045 ** | 0.175 *** | -0.095 | -0.235 | -0.114 |
| | (0.022) | (0.020) | (0.022) | (0.064) | (0.131) | (0.173) | (0.194) |
| HDHP | -0.064 ** | -0.104 *** | -0.085 *** | -0.059 | -0.041 | -0.013 | 0.882 |
| | (0.029) | (0.026) | (0.029) | (0.070) | (0.329) | (0.541) | (0.869) |
| Log(coding intensity) | Coef | Coef | Coef. | Coef. | Coef. | Coef. | Coef. |
| | (Std. Err.) | (Std. Err.) | (Std. Err.) | (Std. Err.) | (Std. Err.) | (Std. Err.) | (Std. Err.) |
| EPO | -0.035 | 0.007 | -0.033 * | 0.002 | 0.128 | -0.139 | 0.033 |
| | (0.035) | (0.018) | (0.019) | (0.031) | (0.133) | (0.132) | (0.093) |
| HMO | 0.028 * | -0.006 | -0.004 | -0.055 *** | 0.035 | 0.084 | -0.116 |
| | (0.015) | (0.008) | (0.008) | (0.016) | (0.180) | (0.187) | (0.083) |
| POS | -0.011 | -0.018 ** | -0.012 | -0.039 ** | 0.008 | -0.115 | -0.004 |
| | (0.016) | (0.009) | (0.009) | (0.018) | (0.100) | (0.101) | (0.072) |
| COMP | -0.012 | 0.041 *** | 0.044 *** | 0.013 | 0.051 | 0.133 | -0.188 |
| | (0.016) | (0.008) | (0.009) | (0.018) | (0.166) | (0.187) | (0.132) |
| CDHP | -0.002 | -0.021 ** | -0.025 *** | -0.006 | 0.113 | -0.139 | -0.103 * |
| | (0.017) | (0.009) | (0.009) | (0.021) | (0.095) | (0.089) | (0.061) |
| HDHP | -0.032 | -0.032 *** | -0.031 ** | -0.046 ** | 0.041 | 0.072 | -0.162 |
| | (0.022) | (0.011) | (0.012) | (0.023) | (0.241) | (0.297) | (0.267) |

Notes: All results use N=20,430,383 spells in which payments for procedures are positive. Results from 28 regressions are presented. For each model a preliminary regression was run at the individual level. Coefficients on each of the coefficient of the (plantype*employer*year*single) combinations are used in a second stage regression, with standard errors as shown in parentheses here. Fitted probabilities from 222 preliminary logit models estimated for each employer*year*single used as instruments for TSLS models. Ln(real payments for procedures| >0) = actual payments deflated by a Medicare's Geographic Price Cost Index (practice costs) and year. Ln(pricing intensity) = Ln(Spending using actual prices/Spending using US average prices). Ln(treatment intensity) =Ln(Spending using US average prices/Expected spending | concurrent RRSt). Ln(coding intensity) =ln((Expected spending |concurrent RRSt)/(Expected spending |RRSt-1 and coarse concurrent diagnoses)).

# Appendix A: Additional tables

Table A-1: Health Plan Types

| Plan Type | Acronym | Description |
|---|---|---|
| Exclusive Provider Organization | EPO | Patients must get care from in-network providers; must choose PCP; no capitation |
| Health Maintenance Organization | HMO | Same as EPO, but at least some services are paid on capitated basis |
| Point-of-Service, noncapitated | POS | Patients have financial incentives to use in-network providers; PCP referrals necessary |
| Point-of-Service, Full or Partial Capitation | POSC | Same as POS, but at least some services are paid on capitated basis |
| Preferred Provider Organization | PPO | Patients have financial incentives to use in-network providers; no "gate-keeping" |
| Comprehensive | Comp | No incentive for patient to use particular providers |
| Consumer-Driven Health Plan | CDHP | PPO plan with a Health Reimbursement Arrangement |
| High Deductible Health Plan | HDHP | Statutory High Deductible Health Plan coupled with a Health Savings Account |

Note: PCP = Primary Care Practitioner. Plan type definitions based on 2011 Truven MarketScan Commercial Claims and Encounters Database User's Guide.

Table A-2 Descriptive statistics in two estimation samples

| | Full 5M sample | | Employer subsample | |
|---|---|---|---|---|
| | Mean | Std Dev | Mean | Std Dev |
| Age of contract holder | 45.99 | (9.77) | 46.30 | (9.53) |
| Male (%) | 55.34 | (49.71) | 51.92 | (49.96) |
| Annual concurrent RRS250 for 12 months prior to current month | 1.02 | (1.74) | 1.01 | (1.65) |
| Annual prospective RRS250 for 12 months prior to current month | 1.21 | (1.31) | 1.21 | (1.25) |
| Nominal Payment | 359.76 | (3778.55) | 339.54 | (3276.14) |
| Real Payment | 383.26 | (3872.46) | 369.63 | (3516.77) |
| Real Payment for Procedures | 229.26 | (1564.98) | 207.90 | (1134.33) |
| N(person Spell-months) | 251,853,264 | | 62,971,626 | |

Notes: Samples are from MarketScan Commercial Claims and Encounter Data. The full 5M sample selects those who are: adults age 21 to 64, enrolled for 60 months, 2007-2011, non-missing age, gender, region, have at least one month of plan type information in any year. The employer subsample is after further selecting enrollees to whom we can assign an employer, health plan type, and PCP.

Table A-3 Alternative risk adjustment models and their R-squares predicting monthly treatment spell spending on procedures when spending is positive

| | Diagnostic information used: (1) Prospective | (2) Concurrent | (3) Both prospective and concurrent |
|---|---|---|---|
| Coarse information (RCCs) | $R^2$ = 0.0394<br>$R^2$ = 0.0456<br>d.f. = 167 | $R^2$ = 0.2015<br>$R^2$ = 0.3177<br>d.f. = 167 | $R^2$ = 0.2025<br>$R^2$ = 0.3201<br>d.f. = 279 |
| Rich information (HCCs) | $R^2$ = 0.0488<br>$R^2$ = 0.0420<br>d.f. = 58 | $R^2$ = 0.1927<br>$R^2$ = 0.2551<br>d.f. = 66 | $R^2$ = 0.2296<br>$R^2$ = 0.3434<br>d.f. = 282 |

Notes: All results were generated using the same sample of 68,496,716 treatment spell months with positive spending on procedures. Each cell contains the $R^2$ and degrees of freedom used in two different regressions. The first $R^2$ is uses spending on procedures in their natural form, while the second uses the log of spending.

All regressions included 12 age-sex dummies and 47 spell month dummies. Coarse information models in the first row use only RCCs, while models using rich information in the second row use HCCs as captured in a single Relative Risk Score (RRS) based on HCCs as captured in the DxCG models for topcoded medical spending.

Table A-4 Alternative models of health plan type choice at contract level (N=6,162,116)

| Model # | Model description | RHS variable count | Log Likelihood function (LLF) | Incremental LLF relative to base model | $R^2$ |
|---|---|---|---|---|---|
| Base | Base model | 7 | -3577656 | 0 | 0.3965 |
| A-1 | employee age | 6 | -3611247 | -33592 | 0.3922 |
| A-2 | employee sex | 6 | -3592742 | -15086 | 0.3940 |
| A-3 | family size | 6 | -3582391 | -4735 | 0.3957 |
| A-4 | spouse | 6 | -3590507 | -12852 | 0.3945 |
| A-5 | baby flag | 6 | -3578030 | -374 | 0.3964 |
| A-6 | sum of adult prospective $RRS_{-1}$, topcoded | | -3595774 | -18118 | 0.3933 |
| **Alternative Risk Scores** | | | | | |
| Base | Using sum of all-adult prospective $RRS_{-1}$, topcoded at 95% | 7 | -3577656 | 0 | 0.3965 |
| A-7 | Using employee-only concurrent $RRS_{-1}$, un-topcoded | 7 | -3589143 | -11488 | 0.3944 |
| A-8 | Using employee-only concurrent $RRS_{-1}$, topcoded at 95% | 7 | -3583757 | -6101 | 0.3951 |
| A-9 | Using employee-only prospective $RRS_{-1}$, topcoded at 95% | 7 | -3578843 | -1188 | 0.3961 |
| A-10 | Using sum of all-adult concurrent $RRS_{-1}$, topcoded at 95% | 7 | -3582467 | -4812 | 0.3955 |
| A-11 | Using sum of all-adult prospective $RRS_{-1}$, untopcoded | 7 | -3584188 | -6532 | 0.3953 |
| A-12 | Using employee-only concurrent $RRS_{-1}$ topcoded at 95% AND sum of all-adult prospective $RRS_{-1}$, topcoded at 95% | 8 | -3573292 | 4364 | 0.3973 |
| **Model using only plan constants** | | | | | |
| A-13 | Using plan type dummies only | 1 | -3696457 | -118801 | 0.3790 |

Notes: Each row is from a different run estimating 121 logit models on 6,162,116 employees at 43 employers offering 402 distinct plan type offerings in each of four years from 2008 to 2011. This sample excludes people without an identifiable employer, or at an employer offering only one plan type, or any plans with fewer than 30 enrollees. All relative risk scores (RRS) were calculated using Verisk Health/DxCG software that used diagnostic age, and sex information from the year prior to the plan choice modeled. Four different types of RRS are considered: Concurrent models (predicting same year spending) and prospective models (predicting subsequent year spending), for the employee only and for the sum of all adults in the household. Two versions of each of these four models were considered: untopcoded and topcoded. Since untopcoded models can have large outliers that distort logit estimates, preferred specification all use topcoded RRS, topcoded at the 95% for that model variant.
The base model contains the right hand side (RHS) variables prior-year employee age, sex, family size,
The first series of models drop one explanatory variable at a time to calculate the incremental log
The second series of models examines alternative RRS, using alternative concurrent and prospective
The log-likelihood funciton (LLF) column is the sum of all model LLF for each of the 221 employer-year
Incremental LLF is calculated relative to the base model.
The $R^2$ reported is from a regression of the actual choices on the fitted probabilities of each choice,
The final model A-13 is a reminder that a simple model in which plan choices are predicted using only plan type dummies for each employer also does surprisingly well. Most of the variation is due to employer variation in plan types offered, not employee characteristics.

Table A-5: Spell Transition Probabilities

| | Period t spell types: | | |
|---|---|---|---|
| Period t-1 spell types: | No Visit (N) | New Visit (V) | Continuation (C) |
| No Visit ($N_{-1}$) | $1 - Pr(V\|N_{-1})$ | $Pr(V\|N_{-1})$ | 0 |
| New Visit ($V_{-1}$) | $1 - Pr(C\|V_{-1})$ | 0 | $Pr(C\|V_{-1})$ |
| Continuation ($C_{-1}$) | $1 - Pr(C\|C_{-1})$ | 0 | $Pr(C\|C_{-1})$ |
| All | Pr(N) | Pr(V) | Pr(C) |

Notes: Lagged information is indicated by -1, with N, V, and C denoting no visit, new visit, and continuation spells, respectively.

Table A-6 Regression results showing health plan effects on unconditional and conditional visits

| | Pr(any visit) | $Pr(V\|N_{-1})$ | $Pr(C\|V_{-1})$ | $Pr(C\|C_{-1})$ |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Coef | Coef | Coef. | Coef. |
| | (Std. Err.) | (Std. Err.) | (Std. Err.) | (Std. Err.) |
| EPO | -0.030 ** | -0.013 | 0.016 | -0.038 ** |
| | (0.01) | (0.01) | (0.02) | (0.02) |
| HMO | -0.048 *** | -0.030 *** | -0.027 *** | -0.060 *** |
| | (0.01) | (0.00) | (0.01) | (0.01) |
| POS | -0.013 ** | -0.006 | -0.009 | -0.018 ** |
| | (0.01) | (0.00) | (0.01) | (0.01) |
| COMP | -0.048 *** | -0.042 *** | -0.023 ** | -0.013 |
| | (0.01) | (0.00) | (0.01) | (0.01) |
| CDHP | -0.044 *** | -0.024 *** | -0.027 *** | -0.032 *** |
| | (0.01) | (0.00) | (0.01) | (0.01) |
| HDHP | -0.077 *** | -0.042 *** | -0.062 *** | -0.070 *** |
| | (0.01) | (0.01) | (0.01) | (0.01) |

Notes: Results used grouped data to show the effects of health plan on probability of seeking care. Spells were first summarized to the plan type*employer*year*family/single coverage and then used for grouped regressions. All the models control for age sex groups, employer*family, spell month time dummies, and HCC based RRS from diseases recorded during the previous 12 months. Column (1) corresponds to the probability of any visit, regardless of new or continuation visits. Column (2) is the probability of new visit conditional no visit in previous month. Column (3) is the probability of continuing visit conditional on no visit in previous month. Column (4) denotes the probability of continuing visit conditional on a previous continuation spell month of visit.

## Appendix B: Data cleaning steps

MarketScan (MS) commercial claims data has two kinds of claims, those originating from employers and those originating from health plans. Health plan identifiers ("plan key"), which also identify employers, are only provided on a subset of the employer sample, and only for a subset of each employer's years. Since a key feature of the MS employer data is that when an enrollee changes employers, his enrollee ID will change, we can reliably infer that if a person is in the sample for all five years, then that household is almost always associated with the same employer (or affiliate if there is a merger or breakup). Hence for each household in the employer sample we assigned the employer from a previous or subsequent year. This roughly quadrupled our sample with an employer assigned.

As summarized in Appendix table A1, the MS data contains a diverse set of different health plan types that we wish to include. Health plan identifiers ("plan key") that index a specific plan are missing on 75.9% percent of the eligibility records, and is only present for all five years on 5.3% of all enrollees. However health plan type (as shown in Table 1) is provided for over 98 percent of all enrollees, and is present for all five years on 97%. We conducted an analysis of plan IDs and plan types and determined that where plankey is present, employers on average offer 2.03 plankey ID for each given plan type offered that year. Employers are more likely to offer multiple PPOS, HMOs and POS, the three most common plan types, and less likely to offer multiple EPO, CDHP or HDHP plans if they offer these plan types at all. Despite the limitations, we conduct our analysis at the employer-year-plan type level instead of the unique plan key level. While this blurs the distinction between different HMO, PPO, POS plan variants, it preserves a large sample size, and the distortion on the three new plan types – EPO, CDHP and HDHP is small.

MS claims contain flags when capitation payments within a plan are made, which means that payments are likely understated. We excluded all spells in which we detect any payments based on capitation. This includes all individuals in a plan type called "Point of Service with Capitation" (0.115 percent of the 5.1 million base sample), and spells containing a flag indicating a capitation payment (mostly in PPO and POS plans). Eligibility records with missing enrollee id, age, sex, or plan type were also excluded. Benefit plan features of the MS plans were not available to us. Therefore we adopted an empirical approach that does not require copayment, coinsurance or deductible information.

Employee county is missing on a significant number of enrollees and missing on all claims for 2011 data. Although we developed an algorithm for assigning enrollee counties to claims based on prior year information, our final results were largely invariant to whether county dummies were included or not.

Provider county is missing in a patter similar to enrollee county, and not present in 2011 at all. Since we are using individual and PCP fixed effects, this will largely absorb the provider county. However we also do a geographic adjustment for cost of living at the claims level. For this, we

used the provider county where available, and if not, then the state in which the enrollee lives. On the subset where both the provider county and enrollee state were present, the correlation between price index assigned using the two sources was greater than .99.

Provider specialty is present on 94 percent of all claims on which there is a procedure code. However provider IDs (which can also reflect provider billing numbers) are present only on 46% of such claims. We need both fields to assign PCPs. We explored the data and also found that many provider IDs contain claims for multiple provider specialty, no doubt reflecting primarily multispecialty clinics, hospitals, and other facilities. Rather than dropping all such provider IDs from our analysis, we created a new provider ID, which concatenated the provider ID and provider specialty and used that to define a distinct provider. We are aware that this groups together multiple physicians working at the same practice or billing number. Despite the limitations we used this approach, and also conduct sensitivity analysis in an appendix of the implications of only using provider IDs that are uniquely assigned to one specialty. Hence while we did not impute any Provider IDs, we did split up provider IDs into multiple providers when they had multiple provider specialties.

### PCP assignment.

Patients were assigned to PCPs using the algorithm of Song et al (2012), as also implemented in Vats, Ellis and Ash (2013). This algorithm assigns patients to actual providers seen using provider IDs, provider specialty, procedure codes and type of service categories. One difference is that since we were using monthly rather than annual PCP periods, there is less of a need for multiple tie breaking rules. Initially the only provider specialties allowed to be assigned as PCPs were internal medicine, family practice, pediatrics, geriatrics, obstetrics/gynecology, and medical doctor (not elsewhere classified). Upon inspection, it was found that a significant number of preventive visits in an office setting were made to the provider specialty "multispecialty practice" so we chose to include that as an allowed provider specialty for being assigned to be a PCP when none of the above were also seen. A unique provider was identified using the combination of provider ID and specialty, so that services from a specialist or lab at a given provider ID were not counted in assigning PCPs. Priority was given to assigning PCPs using the first PCP in a spell to provide a preventive visit, but if no preventive visits, then the first PCP to provide an office visit, and if not, then the first PCP to provide any service to a patient. If multiple PCPs of a given quality match were observed for a patient in a given month, then the patient was assigned to the first PCP ID visited.

Once a patient is assigned to a PCP for a spell month, then that patient continues to be assigned to that PCP in subsequent spell months unless a different PCP is visited. Consistent with a gatekeeper model, even inpatient stays, emergency room (ER) visits and other forms of continuation treatment continue to be assigned to the identified PCP even with no further contact. Using analogous logic, at the beginning of the sample period for each patient, months with no visits or assigned visits were assigned to the first PCP assigned to a patient. Hence a

patient visiting the ER in January, and then the PCP the following month would have January also assigned to that PCP. Overall switching rates between different PCPs were rare, occurring in only 4% percent of all spell months. Forty one percent of treatment spells with positive spending included at least one visit to a PCP.

## Price Deflation

Real prices in 2011 dollars were created by inflating spending into 2011 so as to level average prices over time. County level adjustments were made separately for procedures and facility claims, using Medicare's practice expense component of the Geographic Practice Cost index (GPCI). Real payments were calculated using these time- and cost-of-living- adjusted prices. We also calculated the US average price for each procedure and type of service (if no procedure) and used these averages to calculate what payments would have been if US average prices had been used instead of the actual prices. The ratio of the real price to the US prices, weighted by quantities of services in a spell is our measure of pricing intensity.

In order to distinguish price variation that is explained by provider choice from price variation due to time trends and cost of living, prices on each claim were first deflated by an inflation adjustment which normalized the average price to be constant in all years, and then normalized prices by dividing by the county or state prices as captured by the Medicare programs Geographic Practice Cost Index (GPCI) as used in the physician time component of the Medicare Part B payments. The practice expense GPCI reflects regional differences in the wages of employees in physician practices, such as nurses and office staff, and differences in median residential rents, which serve as a proxy for office rent. Quantities of services on each claim were multiplied by these real prices to calculate the value of real payments.

Calculation of real prices was done separately for each procedure code, and also for classes of non-procedure claim (e.g. facility charges) so that the national average prices could be multiplied by actual quantities to generate a US average payment that would correspond to that service. The pricing intensity of a spell is the real price for the set of services provided, divided by what payment would have been if the US average prices instead of the providers own prices are used. Both payments are calculated using the actual quantities of services received. Since the units of measuring facility services (hospital days, DRGs, time receiving tests, days of care, hours in ICU, etc.) are not as readily captured as counts of procedures, there is reason to believe the US average payments for procedures are a more accurate payment measure than US average payments for all procedures and facility care. Hence our preferred dependent variable focuses on real payments for procedures, rather than total real payments including both procedures and facility payments.