NBER WORKING PAPER SERIES

ASSESSING INCENTIVES FOR ADVERSE SELECTION IN HEALTH PLAN PAYMENT SYSTEMS

Timothy J. Layton
Randall P. Ellis
Thomas G. McGuire

Working Paper 21531
http://www.nber.org/papers/w21531

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2015

Research for this paper was supported by the National Institute of Mental Health (R01 MH094290) and the Laura and John Arnold Foundation. The views expressed here are the authors' own and not necessarily those of the Foundation's officers, directors or staff. In addition, Layton was supported by NIMH T32 019733. We are grateful to Konstantin Beck, David Cutler, Mike Geruso, Laura Hatfield, Lukas Kauer, Albert Ma, Joe Newhouse, Sherri Rose, Christian Schmid, Mark Shepard, Richard van Kleef, Wenjia Zhu, and participants at the BU-Harvard-MIT Seminar and the NBER Summer Institute for comments on a previous version. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Assessing Incentives for Adverse Selection in Health Plan Payment Systems
Timothy J. Layton, Randall P. Ellis, and Thomas G. McGuire
NBER Working Paper No. 21531
September 2015
JEL No. I11,I13,I18

ABSTRACT

Health insurance markets face two forms of adverse selection problems. On the demand side, adverse selection leads to plan price distortions and inefficient sorting of consumers across health plans. On the supply side, adverse selection creates incentives for plans to inefficiently distort benefits to attract profitable enrollees. These problems can be addressed by features of health plan payment systems such as reinsurance, risk adjustment, and premium categories. In this paper, we develop Harberger-type measures of the efficiency consequences of price and benefit distortions under a given payment system. Our measures are valid, that is, based on explicit economic models of adverse selection. Our measures are complete, in that they are able to incorporate multiple features of plan payment systems. Finally, they are practical, in that they are based on the ex ante data available to regulators and researchers during the design phase of payment system development, prior to observing actual insurer and consumer behavior. After developing the measures, we illustrate their use by comparing the performance of the payment system planned for implementation in the ACA Marketplaces in 2017 to several policy alternatives. We show that, in protecting against both types of selection problems, a payment system that incorporates reinsurance and prospective risk adjustment out-performs the planned payment system which includes only concurrent risk adjustment.

Timothy J. Layton
Department of Health Care Policy
Harvard Medical School
180 Longwood Avenue
Boston, MA 02115
layton@hcp.med.harvard.edu

Randall P. Ellis
Department of Economics
Boston University
270 Bay State Road
Boston, Mass 02215
ellisrp@bu.edu

Thomas G. McGuire
Department of Health Care Policy
Harvard Medical School
180 Longwood Avenue
Boston, MA 02115
and NBER
mcguire@hcp.med.harvard.edu

1. Introduction

Public policy in the U.S. relies on private individual health insurance markets to provide coverage to about one third of the elderly in Medicare and to all those eligible for the new Marketplaces (formerly called Exchanges) created by the Affordable Care Act (ACA). As is well-known, individual health insurance markets are vulnerable to adverse selection, the tendency of sicker, higher-cost consumers to choose more generous coverage.[1] This natural pattern of demand causes two problems: 1) equilibrium premiums reflect selection as well as coverage differences, leading to price distortions that cause consumers to choose the "wrong" plans (Einav and Finkelstein, 2011), and 2) insurers distort the coverage of their health plans to make them less attractive to sicker enrollees (Glazer and McGuire, 2000). Economic analysis contends with these selection-related inefficiencies by *ex ante* study of the incentives embedded in insurance markets under alternative policy environments[2] and by *ex post* evaluation of the performance of implemented policies based on actual consumer and insurer behavior.[3]

This paper develops and implements a general methodology for assessing the *ex ante* selection-related incentives created by a health plan payment system. Our perspective is at the market design phase: with data on patterns of utilization representative of the population to be covered, we want to develop an approach to assessing how well a payment system – meaning the set of policies regulating both the premium structure and the plan payment scheme – will contend with selection problems. *Ex ante* analysis and simulations are the primary way regulators evaluate and decide on payment systems for the new Marketplaces, Medicare's payment system for private health plans, and plan payment systems in the Netherlands, Switzerland, Israel and Germany.[4] Studies in this literature focus almost entirely on just one component of health plan payment systems: the risk adjustment scheme, using statistical measures such as the R-squared of a regression of actual costs on the predicted costs output by the risk adjustment formula. Some papers in this literature also

---

1 Table 9 in Cutler and Zeckhauser (2000) summarizes thirty studies documenting adverse selection in health insurance. For a more recent review emphasizing studies relevant to the state-level Marketplaces, see McGuire et al. (2014).

2 An example we discuss later is the *ex ante* evaluation of the risk adjustment system used to pay plans in the new Marketplaces by federal contractors (Kautter et al., 2014).

3 See Erickson and Starc (2012) for an early evaluation of insurance pricing in the precursor to national reform in Massachusetts, and Kowalski (2014) for an evaluation of selection inefficiencies in plan choice in state health care reforms. We discuss a number of these *ex post* evaluation studies below.

4 See as examples, Kautter et al., (2014) on US Marketplaces; Pope, et al., (2011) on US Medicare; Shmueli, et al. (2010) on Israel; Beck, Trottman and Zweifel (2010) on Switzerland; Breyer, Heineck and Lorenz (2003) on Germany; Van Kleef, Van Vliet and Van de Ven (2013) on the Netherlands.

include "predictive ratios," the ratio of predicted costs to actual costs for selected groups in the population, such as those with a chronic illness.[5] Although widely used, neither an R-squared nor a predictive ratio has been shown to have a direct interpretation in terms of economic inefficiency.

We derive alternative measures for *ex ante* evaluation of payment system performance, requiring three things from the measures we propose: First, they should be *valid*, i.e., the measures should follow from formal analysis of the economic behavior causing the selection problems the payment system is designed to correct and they should be based on the effects of a payment system on consumer welfare rather than just the size of the distortion created by the selection problem. Second, measures should be *complete* in the sense of accommodating all relevant features of payment systems used to pay health plans including multiple premium categories and reinsurance, not just the risk adjustment formula. Third, the measures should be *practical*, that is, readily computable from the large claims databases available at the design phase for *ex ante* evaluation of payment system alternatives. We *do not* require that the measures evaluate payment system performance using a standard metric such as the total dollars of welfare loss. Instead, we only require that the measures we develop allow for the *comparison* of payment systems according to their implications for social welfare, providing policymakers with a basis for determining which payment systems are "better" than others.

While the standard measure used for *ex ante* analysis of payment system performance, the R-squared from a regression of costs on risk adjustor variables, is both practical (requirement 3) and allows for the comparison of payment system performance, it does not accommodate the complete set of payment system features (requirement 2), nor has it been shown to measure any parameter of *economic* importance (requirement 1). The same could be said of predictive ratios. Other more economically valid welfare metrics, such as the certainty equivalent measure used by Einav et al. (2010) and others, require estimates of key behavioral parameters among the population of interest, causing them to lack the practicality (or feasibility) of the more popular R-squared and predictive ratio metrics and to not be used in actual payment system design. In a nutshell, our paper intends to equip researchers and regulators with a methodology, or a "toolkit," for evaluating payment system alternatives to replace the R-squared/predictive ratio analyses that have been conducted in this literature for more than 30 years.

---

[5] European researchers more commonly study over- and undercompensation – the difference between projected revenues and cost rather than their ratio. See Van Kleef et al. (2013).

We address selection-related efficiency problems associated with both inefficient plan choice and inefficient plan design. In each case, we start with an economic description of the behaviors associated with adverse selection, and derive a measure (or measures) of efficiency loss due to the incentive problems. With respect to inefficient plan choice, we show that a combination of two measures, which we call "premium fit" and "payment system fit," characterize the magnitude of the efficiency loss. Premium fit measures how well premium categories explain the variation in spending in the population, while payment system fit measures how well simulated plan revenues for an individual, which are a function of payment system features such as premium variation, risk adjustment, and reinsurance, match that individual's total cost to the insurer. These measures incorporate the important insight that even if plans charge the best single premium, except in a very special case, some consumers will still sort inefficiently (thereby linking to papers by Bundorf, Levin, and Mahoney (2012), Geruso (2014), and others). With respect to inefficient plan design, we derive a measure similar to a grouped R-squared that recognizes that plans may distort service offerings to avoid unprofitable groups where groups are defined based on individual characteristics, by patterns of service utilization, or by some combination of these two dimensions. All measures can be readily calculated with data typically available at the *ex ante* stage of payment system design, without estimates of any difficult- or impossible-to-estimate behavioral parameters. We believe our proposed measures satisfy our criteria of practicality, validity and completeness.[6] We provide simple expressions for these measures to allow for easy implementation by policymakers.

We illustrate the use of our measures for the evaluation of the performance of the payment system proposed for implementation in the ACA Marketplaces starting in 2017 (ACA 2017) relative to other counterfactual payment systems. Under ACA 2017, premiums will continue to be set by a pre-specified age curve, risk adjustment will continue to be concurrent (i.e. based on diagnoses from the current year rather than the prior year) and mandatory federal reinsurance will be discontinued.

[6] We recognize that our three selection-related measures do not constitute a comprehensive evaluation of plan payment systems. Regulators have other objectives, including incentives for cost control, avoiding gameability, and fairness. For example, Handel, Hendel and Whinston (2015) show an explicit tradeoff between the efficiency properties we study and "reclassification risk." Similarly, Beck et al. (2014) show that there is an implicit tradeoff between the efficiency properties we study and a concept of "fairness." Additionally, Geruso and McGuire (2015) show a similar tradeoff between selection inefficiencies and incentives for cost control. To maintain simplicity and tractability, we will have little to say about these other, important objectives. Instead, we focus on improvements over the selection-motivated R-squared and predictive ratio measures (which also do not assess cost control, reclassification risk, fairness or gameability). We return in a final section to discuss other dimensions of comprehensive *ex ante* payment plan evaluation.

We evaluate three alternative payment systems, cumulatively altering the three main features of the ACA 2017 proposal: 1) switching to prospective (i.e. based on diagnoses from the prior year) rather than concurrent risk adjustment, 2) keeping reinsurance, and 3) freeing premium age categories to be set in market equilibrium rather than by a pre-specified age curve. These simulations illustrate how our selection metrics work, and also provide relevant evidence for policy choices going forward for the Marketplaces. As we explain in detail below, the data we use to estimate the measures are an updated (and better selected) version of the data that was used to calibrate (and evaluate) the current federally recommended risk adjustment models, so we believe our analyses are informative for policy decisions.

We find that the ACA 2017 payment system is out-performed on all measures by a payment system incorporating prospective risk adjustment combined with the reinsurance policy used in the Marketplaces in 2014. Our results suggest that it is feasible to move away from the concurrent risk adjustment system forced on the Marketplaces in 2014 by the absence of a "history" on plan enrollees to a prospective risk adjustment system without exacerbating problems caused by adverse selection.[7] We also show that when considering sorting inefficiencies the regulated age curve hurts efficiency by forcing plans to charge prices set by the (slightly inaccurate) curve rather than by the market, though the difference is small in case of the ACA curve. Finally, we test the sensitivity of two of our measures to various assumptions about what portion of medical spending is "predictable," a concept that our models suggest is relevant for those measures. Among the payment systems we study, we find that results are largely robust to a fairly comprehensive set of alternative specifications of "predictable costs."

The remainder of the paper proceeds as follows: Section 2 reviews and critiques existing methods for assessing payment systems and measuring selection inefficiency and indicates the gaps in the literature that we seek to fill. We next present theoretically grounded and practical measures related to inefficiency in consumer choice of plan (Section 3) and in insurer choice of the form of the insurance product (Section 4). After a description of our data and empirical methods in Section 5, Section 6 reports our metrics for a series of alternative health plan payment systems for Marketplaces in 2017 and some robustness checks. We summarize and discuss the findings in Section 7.

[7] Not all enrollees in 2017 will have a history. We deal with this in our simulations by assuming that in the prospective risk adjustment case, only age and gender can be used for the large share of enrollees (50%) without a history.

2. Assessing Adverse Selection in Health Insurance Markets

The literature assessing adverse selection falls into three groups. The first and largest group contains papers applying statistical measures to assess payment systems. Papers in the second and third groups apply measures based on the economics of consumer and plan behavior, respectively.

2.1 Statistical Fit of a Risk Adjustment Formula

The literature applying statistical measures of fit evaluates the risk adjustment component (only) of plan payment. Here we discuss the development and evaluation of the risk adjustment system used in Marketplaces as a relevant and representative example of this literature. Kautter et al. (2014) describe the data, methods and results for the risk adjustment formulas developed for the Department of Health and Human Services (HHS) for use in the state Marketplaces beginning in 2014.[8] The "HHS-HCC" model, with HCC denoting the Hierarchical Condition Categories that comprise the diagnosis-based indicators for chronic conditions used as variables in estimation and payment, is based on an earlier Center for Medicare and Medicaid Services (CMS) version developed for Medicare (the "CMS-HCC" model from Pope et al., 2011). In addition to being modified and calibrated for a younger population with different coverage, the HHS-HCC model is "concurrent" whereas the model for Medicare is "prospective." In 2014, Marketplace enrollees had no observed history of health care claims to use for risk adjustment, forcing the HHS-HCC model to use current-year diagnoses to "predict" annual plan costs from the same year. Data were from a population with employer-sponsored health insurance, and separate models were estimated for different age categories and plan metal levels.[9] The HHS-HCC model is being used in all states with the exception of Massachusetts,[10] and will be the basis of our evaluations reported in Section 6.

[8] Comprehensive review chapters in Volumes 1 and 2 of the *Handbook of Health Economics* deal with risk variation and risk adjustment. Van de Ven and Ellis (2000) cover many of the econometric issues associated with constructing a risk adjustment formula. Breyer, Bundorf and Pauly (2012) discuss the attributes of good risk adjuster variables, and update the use of risk adjustment in health care systems internationally.

[9] See Kautter et al. (2014), pages E6-E8 for explanation and discussion of these issues.

[10] Massachusetts had its own risk adjustment formula for use in its health care reform implemented in 2006. Starting in 2014, Massachusetts has adopted a new risk adjustment model very similar to the HHS-HCC model but estimated on data only from New England states. For more information on the Massachusetts risk adjustment model, see HealthConnector (2013).

After developing the HHS-HCC model, Kautter and colleagues conducted an *ex ante* analysis of the Marketplace payment system, reporting two types of statistics:[11] The first, used to assess explanatory power at the individual level, is the model R-squared, the percent of the total variation in plan liabilities explained by age-gender categories, concurrent HCC indicators, and selected HCC interactions. For the adult sample (approximately 14m observations), the estimated R-squared was between 0.35 and 0.36 (Kautter et al., 2014, E12). Using current-year diagnostic information to predict costs more than doubles the explanatory power in relation to the prospective CMS-HCC model used for Medicare (Pope et al., 2011).[12] R-squared is by far the most common but not the only statistic used to evaluate risk adjustment models in the literature. Arguments for the less-common alternatives to R-squared are generally made on statistical rather than economic grounds.[13]

The statistic used to assess under- and over-prediction for subgroups is the predictive ratio, defined as the ratio of plan liabilities predicted by the HHS-HCC model divided by the actual costs for a subgroup of potential enrollees. The numerator, predicted liabilities, would be the revenues for the subgroup if the risk adjustment formula were the only factor determining plan payments. A predictive ratio near 1.0 indicates for the group in question that the predictive model matches predictions (payments) to actual costs at the group level. Kautter et al. (2014, E22) computed predictive ratios for various subgroups defined by predicted costs.

In other papers and reports, model predictions and actual costs have been compared for subgroups defined in various ways. In their evaluation of the CMS-HCC model, Pope et al. (2011)

---

[11] Prior to discussion of the statistics used for evaluation, the authors first chose the diagnostic variables to include in the model based on clinical considerations. Kautter et al. (2014, E4) identify four criteria applied to select diagnostic categories (the HCCs) to be included in the model from among the 264 possible categories, summarized as: clinical significance, not discretionary or subject to "diagnostic discovery", not indicators of poor quality, and indicators of predictable or chronic conditions (not random acute events). On these bases and after some grouping, 100 HCC factors were selected for inclusion in the model.

[12] A more complete analysis of the factors that account for the difference in explanatory power between the CMS-HCC model and the HHS-HCC model would also consider the fact that while the CMS-HCC model predicts only medical spending, the HHS-HCC model predicts both medical spending and the more predictable spending on prescription drugs. Additionally, the HHS-HCC model is estimated on a younger population for whom, unlike with the older Medicare population on which the CMS-HCC model is estimated, mortality is a rare event, resulting in a smaller variance in annual spending.

[13] The mean absolute prediction error (MAPE) uses a linear rather than quadratic loss function. See, for example, Van Barneveld et al. (2001) and Ettner et al., (2001). Van Veen et al. (2015) summarize fit measures used in this literature, and document that MAPE is the second most popular measure of individual fit. The vast majority of papers use an R-squared statistic (or closely related) measure of fit of the risk adjustment formula and/or predictive ratios with predicted values from the risk adjustment formula in the numerator. For an alternative framework with asymmetric weighting of over and underpayments, see Lorenz (2014).

report predictive ratios for a large number of subgroups, including groups defined by disease, numbers of prior hospitalizations, demographic characteristics, and others. Van Kleef et al. (2013) merged survey information with health claims for a subset of people in the Netherlands to calculate "undercompensation" (defined as the difference in costs and predicted revenue rather than their ratio) for various groups of people, including those with low physical and mental health scores and those with chronic conditions.[14] They compare seven different risk adjustment models with different sets of explanatory variables. Brown et al. (2014) divide FFS Medicare enrollees into groups according to their percentile of *ex post* spending and calculate the average level of undercompensation for each group.

R-squared and predictive ratio measures ignore payment system features other than risk adjustment that affect both the fit of revenues to costs at the individual level as well as under- and overpayment for subgroups. Premiums also determine revenue to plans (as in Medicare Advantage, for example) and affect payment system fit and incentives for both individuals and groups. Other payment system features such as reinsurance also help match payments to costs for individuals and groups. Judging how well the full payment system fits costs requires taking these features into account. Even if the purpose of an analysis is to assess only the risk adjustment methodology, taking account of the other features of payment is necessary to accurately gauge the incremental contribution of risk adjustment.

Simple fixes amend the R-squared and predictive ratio methods to account for other payment system features. Geruso and McGuire (2015) and Layton et al. (forthcoming) construct the fit of the payment system at the individual level by substituting a simulated payment that a plan would receive for enrolling an individual for the regression predicted value. The "payment system fit" measures the "explained variance" in costs accounted for by risk adjustment and reinsurance, not just the variance explained by the risk adjustment model.[15] McGuire et al. (2014) perform a similar modification for predictive ratios. The numerator of the "payment system predictive ratio" for a subgroup is the sum of the payments for the group (which can depend on all payment system features) rather than the regression predicted values. The denominator in these predictive ratio measures remains the actual costs for the groups.

---

[14] The "mean prediction error" for subgroups, as figured in Hsu et al. (2010), is similar.
[15] If r_i is the revenue associated with a person in a payment system and x_i is the person's cost, the "payment system fit" is $\frac{\mathrm{Var}(x_i) - \mathrm{Var}(x_i - r_i)}{\mathrm{Var}(x_i)}$.

More fundamental is that the underlying economic rationale for these statistical measures (even the more comprehensive "payment system measures") is missing. While it is intuitive that better fit in both forms should improve the performance of a payment system with respect to selection problems, we are aware of no rigorous derivation of why these measures are the right ones. One line of critique of R-squared measures argues that "only predictable costs matter" in assessing alternative risk adjustment models. However, "predictability" is not a simple matter to incorporate into metrics of *ex ante* payment system performance.[16] One of the virtues of our approach of deriving metrics from the underlying economic behavior of consumers and plans is that we can see just where predictability (by the social planner, by consumers, or by insurers) figures into the story, allowing us to target robustness analyses around the largely unknown degree to which spending is predictable to the components of the measures where expectations actually matter.

We now turn to two sets of *economic*, rather than statistical, measures related to consumer choice of health plans and insurer choices regarding product design, respectively. After presenting and critiquing the measures, we note whether they can be assessed at the *ex ante* design stage of the health plan payment system.

2.2 Inefficient Consumer Choice of Plan

In Akerlof's (1970) "The Market for Lemons," used cars varied in quality but, because of information asymmetry, were indistinguishable to consumers and traded at one price. Generally, too few used cars were traded, and depending on demand and cost conditions, the market might not exist at all. Cutler and Reber (1998) applied this model to health insurance where enrollees varied in cost (to insure), but because of premium regulation, insurers could only charge one price.[17] Generally, too few consumers buy generous health insurance, and depending on demand and cost conditions, health insurance markets might fall into a "death spiral" and generous plans could disappear altogether. Recent influential papers in this stream of the adverse selection literature are

[16] Some papers propose an empirical measure of "how much of health care costs are predictable" by using extensive sets of information that consumers might have available for prediction, such as five years of past health care spending in Van Barneveld et al (2001) or something similar in Newhouse et al (1989) who estimate individual fixed effects based on several years of data. These predictions may of course under- or overstate how much consumers can actually predict. Researchers then compare the R-squared from a particular risk adjustment formula to this "maximum explainable R-squared."

[17] Premium regulation can cause the same problems as information asymmetries (Pauly, 2008). It may be that information is available to support premium differences by certain characteristics (e.g., gender), but regulation prohibits premium discrimination on this basis. A requirement of "community rating" on premiums prohibits premium discrimination of any kind.

by Einav and Finkelstein and their colleagues.[18] Before turning to the ideas and their application in these papers, we first state the conditions under which consumers choose efficiently among health plan alternatives.

Suppose health plans have fixed characteristics (as they do in the papers noted above). Imagine, for example, consumers choosing between a less generous plan and a more generous plan designated "silver" and "gold," or between plans with different management practices, such as between Traditional Medicare and Medicare Advantage. Assuming consumers must choose some plan, the efficient price or "incremental premium" for a consumer is the difference between the plan cost for that consumer in the more generous plan and her plan cost in the less generous plan (Keeler, Carter and Newhouse, 1998). The argument is the same as that for prices generally: when consumers face prices equal to costs, utility-maximizing consumers make socially efficient choices. Choices are inefficient when consumers do not face the right prices and/or they do not maximize utility.[19] The fundamental adverse selection problem here is from inefficient plan pricing.

Demand-Based ("Sufficient Statistics") Models and a Total Surplus Measure of Welfare Loss

We refer to the Einav and Finkelstein (EF) model as demand-based because the behavior and welfare analyses are based on demand and cost curves (as opposed to the utility functions which form the basis of choice and welfare in other papers).[20] These models are also referred to as "sufficient statistics" models outside of the context of health insurance (see Chetty 2009). In what Einav, Finkelstein and Cullen (EFC, 2010) refer to as their "textbook example," consumers choose between a high-coverage contract, H, and a low-coverage contract, L. The L contract is normalized to be "no insurance" and, hence, is costless and free to all consumers. The (incremental) price of the H contract is denoted by P, which must be the same for all potential enrollees. Consumers purchase the H contract if their valuation (denoted by D(P)) exceeds P (since the price and value of the low contract are both zero). As the premium falls, more enrollees choose H. The characteristics of these enrollees define an average and a marginal cost curve for the H plan. For a given price P, average cost for plan H (denoted by AC(P)) is the average cost of the enrollees who choose to enroll in contract H at that price. "Marginal cost" (denoted by MC(P)) is the average cost of the

---

18 See Einav, Finkelstein and Cullen (2010), Einav, Finkelstein and Levin (2010), Einav and Finkelstein (2011), and Einav et al. (2013).

19 Models from behavioral economics, as well as some of the utility-based papers discussed later in this section consider imperfect utility maximization, such as from "inertia" or "status quo bias."

20 Utility and demand models are clearly connected in theory. Einav, Finkelstein and Levin (2010) develop a utility basis of demand models, and include a discussion of both types and the associated welfare methods.

consumers newly led to buy H when the premium falls to P, i.e., those whose willingness to pay is exactly P. Figure 1 replicates (omitting some labeling) the first figure in EFC (2010).

The (constrained) efficient price (premium) and quantity (enrollment in H) are given by the point where the marginal cost curve intersects the demand curve. The premium that achieves efficient sorting cannot be sustained (without a subsidy) because competition enforces zero profits so that P = AC(P) in equilibrium. The efficiency loss at this P is depicted as the "conventional" welfare triangle, the shaded area in Figure 1.[21] As was acknowledged by EFC (2010), however, if there is preference heterogeneity such that there exist individuals with the same willingness-to-pay but different costs, as will generally be the case, then the welfare triangle in Figure 1 will not fully describe the welfare loss.[22] The reason is that with heterogeneity in the relationship between demand and incremental cost, "marginal cost," MC(P) is, in fact, an *average of the marginal costs* over all individuals whose willingness to pay is exactly P. Thus, even when MC(P) = P, there are individuals (those whose willingness to pay exceeds P but whose cost is higher than their willingness to pay) who join the plan even though, from a social welfare point of view, they shouldn't, and there are individuals (those whose willingness to pay is less than P but whose cost is lower than their willingness to pay) who will choose not to join the plan even though, from a social welfare point of view, they should. A complete welfare analysis of efficiency in consumer choice of plans needs to include welfare losses from a single premium (or, more generally from a limited number of premium categories) as well as inefficiencies from adverse selection and the distortion of the incremental price for the more generous plan it causes.[23] This is especially important when comparing the efficiency properties of payment systems with different premium structures as we do in Section 6 below.

EFC (2010) implement this demand-based welfare framework with data on health plan prices and choices of Alcoa employees along with individual-level data on employee medical

---

[21] There are two cases, the "adverse selection" case, where individuals' willingness to pay for plan H is increasing with their cost (to the plan) and the other, the "advantageous selection" case, where individuals' willingness to pay is decreasing with their cost. In the adverse selection case, the one shown in Figure 1 and the one we work with in this paper, the welfare loss is due to the fact that "too few" individuals join the H plan, relative to the social optimum.

[22] In general, "selection" could occur on many dimensions: healthcare costs, geography, cognitive ability, and other factors with positive or negative relations to expected future health care costs. See footnote 3 in EFC for some discussion. See also Einav and Finkelstein (2011), footnote 6, and Einav, Finkelstein and Levin (2010), page 326.

[23] As Einav, Finkelstein and Levin (2010, p 326) put it: "With richer heterogeneity, we may still be interested in the degree of efficiency that can be realized with a uniform price, but we may also want to understand how the potential for efficient coverage depends on the information available to set prices…"

spending. Prices for health insurance options varied with geography, creating enough price variation for the authors to estimate how demand and cost varied by price. Both demand and cost curves were downward-sloping in price, in accord with the depiction in Figure 1 above.[24] The authors found some evidence for adverse selection into more generous plans (implied by the estimated downward-sloping cost curve) but the estimated welfare loss was very small in absolute terms ($9.55 per employee per year) and very small in relation to the full "area under the demand curve" for the generous plan (less than 3% of that total surplus).

Hackmann, Kolstad and Kowalski (HKK, 2015) use the EF model to evaluate the welfare consequences of the Massachusetts health care reform of 2006, the precursor to the national reform, treating the introduction of the tax penalty as an exogenous price fall for individual health insurance, and finding more substantial welfare gains.[25] Kowalski (2014) applies the HKK version of the EFC model to estimate the welfare consequences of the implementation of the ACA on consumers in the U.S. individual health insurance market, leveraging data on average costs, enrollment and prices in individual health insurance markets through the first half of 2014 for all states except California and New Jersey (which had incomplete data). Kowalski exploits the economy of the EF welfare model, using three "sufficient" statistics, average cost, enrollment, and premiums, and estimates of these in the counterfactual world without the ACA, to estimate the effect of the policy on total surplus.[26]

---

[24] See Einav, Finkelstein and Cullen (2010) Figure V page 914. The cost curves in the figure are not exactly analogous to the curves estimated by EFC (2010). In the figure, individuals choose between insurance and uninsurance, making the incremental average cost curve equal to the average cost curve for contract H. In EFC's empirical setting, however, individuals choose between two insurance plans. In this setting the incremental average cost curve is equal to the difference between contract H and contract L's average cost curves.

[25] The estimated welfare gain was substantial, $335 per person in the individual market. As HKK point out, the movement in Massachusetts was from no insurance to some insurance, so the welfare gain could be expected to be larger than the movement between insurance plan types studied by EFC (2010). In addition, HKK argued that the choice platform in the reform sharpened price competition, reducing insurer markups and generating welfare gains of another $107 per person per year.

[26] HKK (page 1064) make a similar observation, stating their "approach allows us to estimate the welfare impact of reform using available data and a minimum of assumptions about the underlying structural preferences of consumers and competing insurers." As we have indicated and explain later in Section 3, the demand-cost metric is not a complete measure of inefficiency in plan sorting when estimating welfare changes across settings with different premium regulations. This does not necessarily mean that the demand and cost parameters are not "sufficient statistics" for the analysis of the welfare consequences of the policies studied in these papers. Because the Massachusetts reform studied by HKK and all of the alternative counterfactual pricing policies studied by EFC did not involve changes to premium regulations (community rating was required in all cases), these parameters are sufficient for welfare analysis in those settings. However, more generally, when premium regulations change as part of a reform (or as part of a simulated reform), parameters of demand and cost are not sufficient for a full welfare analysis of the reform. For

Utility-based ("Structural") Models

Utility provides an alternative welfare framework to demand-based measures. Rather than estimating a few key parameters or "sufficient statistics," papers using a utility-based, or structural, framework to study the welfare consequences of adverse selection typically estimate the full joint distribution of consumer costs and willingness-to-pay and then use that distribution to simulate consumer sorting and welfare under various premium policies. Utility-based frameworks allow the researcher to perform virtually any kind of counterfactual policy simulation, including policies that alter the permitted premium structure.

Bundorf, Levin, and Mahoney (BLM, 2012) use a utility-based framework to study the welfare consequences of adverse selection. Similar to EFC, BLM study a setting where consumers choose between two insurance plans and different groups of consumers face plausibly exogenous premium variation. However, rather than use premium differences to identify the slope of the demand curve as the application of an EF model would do, BLM use premium differences to estimate the joint distribution of consumer costs and willingness-to-pay for a more comprehensive PPO relative to a more restrictive HMO. This joint distribution of risk and preferences becomes the basis for simulations of the effects of different levels of restrictions on premium variation on consumer sorting and welfare.

In addition to estimating the welfare consequences of adverse selection in a setting with a single premium, BLM recognize that no single premium sorts efficiently and show that this conceptual point is empirically important, finding that the best single price only captures one quarter of the potential welfare gains from efficient sorting, illustrating the advantage of utility-based over demand-based models. Einav et al. (2013) use a similar utility-based model to quantify welfare losses from selection on moral hazard. They estimate the joint distribution of incremental willingness-to-pay for a more comprehensive plan and individual-level "moral hazard," or demand response to the marginal price to the consumer of medical care, also allowing for more exotic counterfactual policy simulations than could be performed with the basic demand and cost parameters estimated by EFC (2010).

example, in the context of the national reform, many states adopted community rating for the first time, making the demand and cost parameters used in the EF framework insufficient for a full analysis of welfare changes due to the ACA, as such an analysis would be missing the efficiency consequences of premium restrictions.

Another set of papers, typically lacking exogenous variation in premiums with which to identify willingness-to-pay, rely on assumptions about risk aversion and study consumer choice between plans varying only in cost-sharing rules. Consumer choices are assumed to be a function of plan premiums, risk aversion, and the consumer's distribution of expected out-of-pocket costs in each plan. The researcher specifies each consumer's distribution of expected out-of-pocket costs using a cell-based method dividing the population into groups based on a predictive model of their future medical spending and levels of risk aversion are then identified by assuming that any variation in choices across plans not explained by variation in expected costs is due to risk aversion. The most prominent paper using this method is Handel (2013) which studies the interaction between consumer inertia, or switching costs, and adverse selection, showing that inertia can attenuate selection problems. Geruso (2014) uses a similar method to study the welfare consequences of allowing premium variation by age and sex. Handel, Hendel and Whinston (2015) use it to study the trade-off between adverse selection and reclassification risk in the Marketplaces. Layton (2014) uses this method to study the welfare consequences of risk adjustment in the Marketplaces, where premiums vary by age. Finally, Handel and Kolstad (2015) critique the method by showing that it tends to over-estimate consumer risk aversion by ignoring "information frictions" that cause consumers to make sub-optimal choices.

For our purposes, it is important to note that both the demand-based and utility-based frameworks are typically *ex post,* i.e., they evaluate the efficiency of plan choice after the market happens, and are not intended to guide design at the *ex ante* stage, though they often do include analyses of important counterfactual policies to provide guidance to policymakers about which types of problems are most important and which policies are most likely to ameliorate those problems.[27] Furthermore, the demand-based framework of EF does not accommodate welfare analysis of alternative premium structures, and the utility-based approaches, while in principle providing a more complete welfare assessment, are much more demanding in both the assumptions and the setting from which the data were extracted required for estimating the necessary economic parameters.

2.3 Plan Actions to Attract Good and Deter Bad Health Insurance Risks

The literature reviewed in Section 2.2 assumes that insurance contracts are fixed. In reality, insurers respond to incentives to attract healthy enrollees and avoid sick ones. Rothschild and

[27] For examples of this type of counterfactual simulation see BLM's (2012) analysis of alternative premium policies and EFC's (2010) analysis of various subsidy policies.

Stiglitz (RS, 1976) were the first to model adverse selection and endogenous insurance contracts. In RS, since premiums are valued equally by the good and bad risks, and "coverage" is valued more by the bad risks, plans (inefficiently) reduce both the premium and coverage to attract the good risks. Glazer and McGuire (2000) applied the RS model to markets for managed care health insurance. Even with regulated premiums plans can use "service-level selection" to distort their offerings of different dimensions of coverage (e.g., sicker people value chronic care more) to attract the good risks. In Marketplace plans and other settings, nominal coverage is regulated, preventing service-level selection on dimensions like covered services. However, plans can work around the regulations to create networks and drug formularies favoring/disfavoring certain conditions or impose more or less strict care management techniques across different categories of care. Breyer, Bundorf and Pauly (2012, p 729) refer to these activities as "indirect selection."

The literature on service-level or "supply-side" selection began with studies of the incentives of insurers to distort service-level offerings to attract good risks based on models of health plan profit maximization. Frank, Glazer, and McGuire (FGM, 2000) identify the characteristics of service-level spending that generate incentives for insurers to inefficiently distort service-level benefits. Ellis and McGuire (EM, 2007), in an application to Medicare Advantage, show that when plans are designed to maximize profits, services that are predictive, predictable and exhibit high demand elasticity are rationed tightly. Predictability, the degree to which enrollees can anticipate future use of a service, is a necessary condition for service-level rationing to matter -- if consumers cannot anticipate their use of a service, they cannot be influenced in their plan choices by its selective rationing. Because risk adjustment and other plan payment system features affect the revenue a plan receives for enrolling an individual, these features also affect plan incentives to ration through predictiveness. By transferring payment to individuals with high costs, the correlation between service use and net revenues can be altered, mitigating or in some cases eliminating incentives to distort services (Glazer and McGuire, 2002). McGuire et al. (2013) study incentives for service-level selection in the Marketplaces with a payment system incorporating risk adjustment and find significant incentives for insurers to discriminate against individuals with chronic diseases. Note, however, that none of these papers connect incentives to distort plan benefits to social welfare.

Other papers assess the evidence for service-level distortions without measuring the incentives to engage in service-level selection. Cao and McGuire (2003) in Medicare and Eggleston and Bir (2009) in employer-based insurance find patterns of spending on various services consistent with service-level selection among competing at-risk plans. Ellis, Jiang and Kuo (2013) rank services according to incentives to undersupply them. Consistent with service-level selection, they show that HMO-type plans tend to underspend on predictable and predictive services (in relation to the average) just as the selection index predicts. This pattern of spending is not observed among enrollees in unmanaged plans. Brown et al. (2014) and Newhouse et al. (2015) study how selection into Medicare Advantage changed after the introduction of risk adjustment. Both studies find that after risk adjustment was introduced, Medicare Advantage plans attracted sicker Medicare beneficiaries. Finally, Newhouse et al. (2013) studies whether Medicare Advantage insurers select groups of beneficiaries with high profit margins, finding little evidence that they do.

Recent work confirms that insurers respond to incentives to distort plan benefits to attract the healthy and avoid the sick. Carey (2014a, 2014b) studies how insurers respond to these incentives in the Medicare Part D prescription drug insurance market. Both papers exploit plausibly exogenous variation in the profitability to a plan of different groups of enrollees. Carey (2014a) uses variation due to changes in the availability of high cost prescription drugs to treat different conditions between the time when the Part D risk adjustment model was initially calibrated and the time when plans were designing their benefit packages and competing for enrollees. Carey (2014b) uses variation due to a re-calibration of the risk adjustment formula. Both papers estimate that insurers charge higher copayments for drugs used by groups of enrollees that are less profitable.[28]

Shepard (2015) studies the interaction between adverse selection and hospital network design in the Massachusetts subsidized health insurance exchange, CommCare, finding that consumers who value the inclusion of a local "star" hospital system most highly also have very high costs. He shows that this is partially due to "selection on moral hazard," or the phenomenon that the people whose spending is most affected by having the star hospital system in their plan's network are the ones who have the strongest preferences for a plan including that hospital. Using simulation, he shows that in equilibrium, selection should result in no plan being willing to include the star hospital system in its network under the current pricing structure, providing, as far as we know, the first welfare analysis of this type of "supply-side" selection problem.[29]

[28] See also Kuziemko, Meckel, and Rossin-Slater (2014) for a study of Medicaid managed care plans attempting to attract lower cost births based on the race-ethnicity of the mother.

[29] While Shepard finds that in equilibrium no plan should cover the "star" hospital, he also finds that, in this particular market, this is efficient: On average, consumers do not value the inclusion of the hospital in a plan's network as much as the social cost of including the hospital.

The extensive literature on supply-side selection problems emphasizes measuring *ex ante* plan incentives with respect to benefit manipulation. The incentive measures are group or service-specific and there has been no standardized metric proposed to compare a payment system creating one set of incentives with another payment system creating different incentives. A few more recent papers study *ex post* performance. Missing is an *ex ante* welfare metric that can be used to guide plan payment choice at the design stage.

3. Consumer Choice of Health Plan and Inefficient Sorting

This section develops a measure of the welfare loss from inefficient sorting of consumers between health plans in the simple setting where all consumers choose one of two plans with fixed characteristics, corresponding to the demand and utility-based papers reviewed in Section 2.2 above. Adverse selection and premium regulations drive a wedge between a consumer's first-best price and the price she is charged in equilibrium leading to Harberger-type measures of welfare loss with the loss proportional to the square of the gap between the efficient and the equilibrium prices.

Insurer costs for the same person may differ across the two plans for various reasons, including coverage differences. For tractability, we specify that the difference between the cost to an insurer from enrolling an individual in the more generous ("gold") plan is proportional to the cost of enrolling her in the basic ("silver") plan. Formally, let x_i be the expected cost to an insurer of enrolling person i in the silver plan and let $(1 + \gamma)x_i$ be the expected cost to an insurer of enrolling person i in the gold plan, with $\gamma > 0$. Person i's expected *incremental* cost is defined as the difference between her expected plan cost in the gold plan and her expected plan cost in the silver plan: $(1+\gamma)x_i - x_i = \gamma x_i$. Consumers fall into T types, where all type-t consumers have the same incremental expected cost: $\gamma x_i = \gamma x_t$ for all $i \in t$. The use of "types" captures an important concept: due to preference heterogeneity, individuals with the same expected costs may exhibit different willingness-to-pay for the gold plan. For now, we define type abstractly, but when applying our metrics to data, we make type empirically operational. Finally, let $prem_t^j$ be the premium the insurer charges a type-t individual to enroll in plan j, j = g,s, and let the incremental premium charged to a type-t individual be $prem_t = prem_t^g - prem_t^s$.

All consumers must choose either the gold or the silver plan. Let $n_t(prem_t)$ be the number of type-t consumers who purchase gold given the incremental price, $prem_t$, and let $P_t(n)$ be the "inverse demand," or the price at which n type-t consumers enroll in gold. In this setting, "incremental" welfare for type-t consumers can be described as a function of $prem_t$:

$$W_t(prem_t) = \left(\int_0^{n_t(prem_t)} P_t(n)dn \right) - n_t(prem_t)\gamma x_t$$

Note that this function is maximized at $prem_t^* = \gamma x_t$, when each individual's incremental premium is set equal to her specific expected incremental cost, the efficiency condition highlighted in Keeler, Carter, and Newhouse (1998) and Bundorf, Levin, and Mahoney (2012). Also note that the expected cost, x_t, is from the *ex ante* view of the social planner, i.e. it is the expected cost given the information available prior to consumers' plan choice, also known as the "rational" expectation. This may differ from the consumer's actual expectation of her future cost which plays a role in the consumer's willingness-to-pay, $P_t(n)$, and which may be less accurate than the rational expectation. The reason the rational expectation is used here instead of the consumer's actual expectation is that welfare is based on the difference between valuation and cost, and cost is from society's point of view, not the consumer's. In other words, we measure social welfare, not expected consumer surplus.

The welfare loss for the group of all type-t individuals due to a price different from $prem_t^* = \gamma x_t$ can be expressed as:

$$\Delta W_t(prem_t) = W_t(\gamma x_t) - W_t(prem_t)$$

$$\Delta W_t(prem_t) = \left[\left(\int_0^{n_t(\gamma x_t)} P_t(n)dn \right) - n_t(\gamma x_t)\gamma x_t \right] - \left[\left(\int_0^{n_t(prem_t)} P_t(n)dn \right) - n_t(prem_t)\gamma x_t \right].$$

Taking a 2nd order Taylor series expansion of $\Delta W_t(prem_t)$ around $prem_t = \gamma x_t$, we show in Appendix A that $\Delta W_t(prem_t)$ is approximated by the following expression:

$$\Delta W_t(prem_t) \approx -\frac{1}{2} n_t'(\gamma x_t)(prem_t - \gamma x_t)^2. \tag{1}$$

The second-order approximation is equivalent to assuming that demand ($n_t(prem_t)$) is approximately linear around γx_t, and implies that the welfare loss for type-t consumers due to a distorted price is proportional to the squared difference between the equilibrium price $prem_t$ and the first-best price $prem_t^* = \gamma x_t$. Summing over types, the welfare loss due to distorted prices in the full population is approximated by:

$$\sum_{t=1}^{T} \Delta W_t(prem_t) \approx -\frac{1}{2} \sum_{t=1}^{T} n_t'(\gamma x_t)(prem_t - \gamma x_t)^2$$

Assuming the elasticity of demand[30] for the plan as a function of the incremental premium is the same for all types, i.e., $\varepsilon = \frac{n'_t}{N_t}$, then:[31]

$$\sum_{t=1}^{T} \Delta W_t(\text{prem}_t) \approx -\frac{\varepsilon}{2}\sum_{t=1}^{T} N_t(\text{prem}_t - \gamma x_t)^2$$

Since all individuals of the same type face the same premium and have the same expected costs:

$$\sum_{t=1}^{T} \Delta W_t(\text{prem}_t) \approx -\frac{\varepsilon}{2}\sum_{i=1}^{N} (\text{prem}_i - \gamma x_i)^2 \tag{2}$$

Expression (2) will represent the exact welfare loss in settings where demand is linear (i.e., $n''_t(\gamma x_t) = 0$).[32] If demand is not linear, then Expression (2) will approximate the actual loss, and the accuracy of the approximation will depend on the size of the gap between the incremental premium for person i, prem_i, and the first-best price, γx_i. The extent of any inaccuracy depends on the joint distribution of prices and expected incremental costs. If the linearity assumption fails to hold and there is a lot of variation in expected incremental costs and little variation in prices, there will be types for which the accuracy of (2) would be questionable. The extent of variation in prices depends on the payment system, with more restrictive payment systems generating less variation in prices. Given our assumption that the cost of insuring an individual in the gold plan is $(1 + \gamma)$ times the cost of insuring her in the silver plan, the variation in expected incremental costs is proportional to the variation in *expected* total costs in the population. So, even if demand is non-linear, as long as premiums are reasonably free to vary and/or the variation in expected costs is not too large, Expression (2) should be a good approximation.[33]

3.1 A Health Plan Payment System: Transfers and Premiums

When efficiency is about consumer sorting between plans with fixed characteristics, Expression (2) shows that the pathway by which a payment system can affect welfare is by moving

---

[30] This is not quite an elasticity although we refer to it as such. Instead, it is the slope of the demand curve normalized by the number of type-t individuals. To be a true elasticity, we would need to divide through by the normalized incremental premium.

[31] Note that $n'_t(\gamma x_t)$ (and thus ε) is negative if willingness-to-pay for the gold plan is greater than willingness-to-pay for the silver plan.

[32] The linear demand assumption effectively assumes that there is uniformly distributed heterogeneity in preferences within types. While there is evidence of a broad range of heterogeneity in preferences (Cohen and Einav, 2007) the uniform distribution assumption may not be close for high-cost types.

[33] Note that the amount of variation in expected costs is related to the predictability of spending. If spending is more predictable, the variation in expected costs is larger.

equilibrium prices, $prem_i$, closer to (or further from) first-best prices, γx_i. Different payment systems will result in different equilibrium premiums and thus different levels of efficient sorting.

A health plan payment system consists of transfers and premium regulations. We define the individual-level transfer under payment system p (which might include risk adjustment and other features) as the net payment from the regulator to the insurer associated with person i:[34] $trans_{p,i} = r_{p,i} - \bar{x}$. This structure follows from the typical risk adjustment system that pays insurers the difference between the predicted cost of an enrollee, $r_{p,i}$, and the average cost in the population. We consider transfers that are budget-neutral in aggregate and independent of plan choice, i.e. $\bar{r}_p = \bar{x}$ and $r^j_{p,i} = r^{j'}_{p,i}$.[35] For example, in a system in which transfers are determined only by risk adjustment, the transfer might be equal to the difference between the product of the normalized risk score and the population average cost and the population average cost. Once the transfer is set, we can determine the plan costs net of transfers for each individual under payment system p, which we denote $x_{p,i} = x_i - trans_{p,i} = x_i - (r_{p,i} - \bar{x})$. The plan must cover these net costs with premiums.[36]

Plans can set premiums separately for each of a set, $\Gamma_p = \{1, \dots, a, \dots, A_p\}$, of premium categories, with the p subscript referring to a particular payment system's categorization of individuals for purposes of premium setting. The number of individuals in premium category a is N_a, with $\sum_a N_a = N$. Plans can vary premiums *across* categories but not among individuals *within* a category. Both premiums and transfers cause plan revenues to vary across individuals, with total revenue to plan j for an individual under payment system p being equal to $rev^j_{p,i} = trans_{p,i} + prem^j_a$.

3.2 Decomposing Welfare Loss

Expression (2) for welfare loss contains the demand-response parameter ε $(= \frac{n'_t}{N_t})$, which is unknown *ex ante*. Rather than assume a value, we construct a metric of the welfare loss under payment system p, *relative to* the welfare loss that would occur under a reference or "base" payment

[34] In the empirical application below in Section 5, we implement transfer and premium rules based on those used in the Marketplaces.

[35] This condition of transfers being independent of plan choice does not hold exactly in either the Marketplace or Medicare Advantage markets. In Marketplaces, there are slight differences in the risk adjustment formula across metal levels and in Medicare Advantage the transfer depends on the level of the plan bid. Furthermore, if plans have different coding practices systematically by plan type, transfers will differ according to plan choice (Geruso and Layton 2015).

[36] For economy of notation, we assume here that the insurer's expectation of person i's future spending (which is relevant for plan costs and pricing) is equivalent to the rational expectation of person i's future spending. We relax this restrictive (and unnecessary) assumption in the empirical section.

system that results in the incremental premium being equal to the average incremental cost in the population, $\mathrm{prem_i} = \gamma\bar{x}$ for all i:[37]

$$\frac{\frac{\varepsilon}{2}\sum_{i=1}^{N}(\gamma\bar{x}-\gamma x_i)^2 - \frac{\varepsilon}{2}\sum_{i=1}^{N}\left(\mathrm{prem_{pi}}-\gamma x_i\right)^2}{\frac{\varepsilon}{2}\sum_{i=1}^{N}(\gamma\bar{x}-\gamma x_i)^2} = 1 - \frac{\sum_{i=1}^{N}\left(\mathrm{prem_{pi}}-\gamma x_i\right)^2}{\sum_{i=1}^{N}(\gamma\bar{x}-\gamma x_i)^2} \tag{3}$$

Such a measure allows us to compare payment systems without making assumptions about consumers' demand response to changes in incremental premiums. Instead, we rely on the fact that the payment system has no effect on demand except insofar as it moves the equilibrium incremental premium, $\mathrm{prem_{pi}}$.

Equation (3) can be decomposed (see Appendix A) into two terms corresponding to the two sources of welfare loss associated with inefficient premiums. This decomposition aids both in the operationalization of the expression for welfare loss and in the intuition behind its interpretation.

$$1 - \frac{\sum_{i=1}^{N}\left(\mathrm{prem_{pi}}-\gamma x_i\right)^2}{\sum_{i=1}^{N}(\gamma\bar{x}-\gamma x_i)^2} = 1 - \left[\underbrace{\frac{\sum_{a=1}^{A_p}\sum_{i\in a}(x_i-\bar{x}_a)^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2}}_{\tilde{\delta}} + \underbrace{\frac{\sum_{a=1}^{A_p}\sum_{i\in a}\left(\mathrm{prem_{pi}}-\gamma\bar{x}_a\right)^2}{\gamma^2\sum_{i=1}^{N}(x_i-\bar{x})^2}}_{\tilde{\phi}}\right] \tag{4}$$

The first term in the expression, $\tilde{\delta}$, captures the relative welfare loss due to the deviation of the second-best price that can be charged given the premium restrictions, $\gamma\bar{x}_a$, from the first-best price γx_i. The second component of (4), $\tilde{\phi}$, captures the welfare losses due to any deviation of equilibrium prices $\mathrm{prem_{pi}}$ from the second-best prices. We discuss these two terms in turn.

3.3 Measuring Inefficiency Due to Limited Premium Categories

The first component of Equation (4), $\tilde{\delta}$, is readily computable given *ex ante* data on expected costs. To put this measure in a form with a more familiar interpretation, we denote one minus this measure, $\delta = 1 - \tilde{\delta}$, as "premium fit," since it captures the efficiency loss due to premium regulations. Note that this component is equal to the R-squared from a regression of (expected) costs on a set of indicator variables for each premium category.[38] For a single premium, the

[37] Such a single premium could be generated by a payment system with a single premium category and transfers that perfectly compensate for variation in individual-level spending.

[38] More precisely, the regression used to estimate this measure should produce predicted values equal to the second-best premiums. In the case of premium categories, the correct regression will be of costs on premium categories. In the Marketplaces, the payment system includes a regulated "age curve" that takes a base premium that plans set for a 21 year old and maps that premium to age-specific premiums using a set of age weights derived actuarially. Under such a payment system, all individuals still effectively belong to the same risk pool, but different individuals will be charged different premiums. Because we wish to estimate our measures for the Marketplace payment system, in Appendix B we re-derive our measures for the special case of a regulated age curve. With an age curve, premium fit is the R-squared from a regression of costs on age

numerator and the denominator of $\tilde{\delta}$ are equal, implying that in our base payment system premium fit is equal to 0, the minimum value δ can take. As premium groups "fit" the distribution of incremental costs better, the measure increases, indicating higher efficiency. At minimum, if premiums fully capture incremental cost variation for each individual, premium fit will rise to 1.

3.4 Measuring Inefficiency Due to Adverse Selection

We now address $\widetilde{\Phi}$, the adverse selection component of (4), which characterizes the inefficiency caused by price distortions beyond the distortion caused by limited premium categories. This is the distortion studied in EFC (2010) and related papers. Because $\mathrm{prem}_{\mathrm{pi}}$ is part of $\widetilde{\Phi}$ and unknown *ex ante* we need to make an assumption about how premiums are set. We follow EFC (2010) and Handel, Hendel, and Whinston (2015) and assume competition forces a plan's premium to be equal to the average (net of transfers) cost of those who choose the plan, abstracting from any additional price distortions caused by imperfect competition. In the presence of multiple premium categories, competition will result in a plan charging individuals in a particular premium category the average (net of transfers) cost of the individuals *within that category* who choose the plan.[39] Denoting as $\bar{x}^{g}_{p,a}$ and $\bar{x}^{s}_{p,a}$ the average *silver plan costs* net of transfers of the individuals joining the gold and silver plans, respectively, this implies that the gold and silver premiums for individuals in premium category z will be equal to:[40]

$$\mathrm{prem}^{g}_{p,a} = (1+\gamma)\bar{x}^{g}_{p,a} \qquad \mathrm{prem}^{s}_{p,a} = \bar{x}^{s}_{p,a}$$

The equilibrium incremental premium for each premium category, prem^{e}_{a}, must then be equal to the difference in these average plan net costs:[41]

$$\mathrm{prem}^{e}_{p,a} = \mathrm{prem}^{g}_{p,a} - \mathrm{prem}^{s}_{p,a} = (1+\gamma)\bar{x}^{g}_{p,a} - \bar{x}^{s}_{p,a} \tag{5}$$

We can now plug our expression for premiums from Equation (5) into the expression for $\widetilde{\Phi}$ from Equation (4) to get:

weights. The age-curve variant on payment system fit is similar to payment system fit as defined above with the addition of a scaling factor equal to one over N times the sum of squared relative age weights.

[39] The equilibration process would be different in Marketplaces where the premium categories are age-related and states impose a single age structure. Equilibration in this case moves the entire age-structure up or down. We re-derive all measures for the setting with an age curve in Appendix B.

[40] Note that the plan costs here potentially differ from the costs from Expression (2). Here, the costs are those that are relevant for the premium-setting mechanism, not the expected costs from the social planner's point of view. For now, we abstract from this issue and assume that the costs relevant for premiums are equal to the rational expectation of each individual's cost.

[41] Our analysis is based on the difference between gold and silver prices. We assume that administrative costs are equal for both plans and do not affect the incremental premiums.

$$\tilde{\phi} = \frac{\sum_{a=1}^{A} \sum_{i \in a} \left((1+\gamma)\bar{x}_{p,a}^{g} - \bar{x}_{p,a}^{s} - \gamma \bar{x}_a \right)^2}{\gamma^2 \sum_{i=1}^{N} (x_i - \bar{x})^2}$$

This measure is effectively a weighted sum of the distortion of each premium group's equilibrium price from its second-best price. Each group's distortion depends on the relationship between demand, costs, and transfers among individuals in the premium group. These relationships could vary across premium groups. We abstract from any differences by making regularity assumptions about the form of these relationships. Sufficient assumptions are laid out in detail in Appendix A, but essentially they require two things. First, they require that for each premium group the difference between the average net cost among individuals choosing the gold plan and the average cost among all individuals in the group is a fixed proportion of the incremental average cost for the group, and this proportion does not vary across groups or payment systems. Second, they require that transfers match costs equally well across premium groups and for individuals enrolled in the gold plan and individuals enrolled in the silver plan.

Given the regularity assumptions, we can re-write $\tilde{\phi}$ as

$$\tilde{\phi} = \omega \left[\frac{\sum_{i=1}^{N} \left(x_i - \left(r_{p,i} + (\bar{x}_a - \bar{r}_{p,a}) \right) \right)^2}{\sum_{i=1}^{N} (x_i - \bar{x})^2} \right]$$

where ω is a scaling parameter that depends on γ and properties of the variation in costs but is independent of the payment system. Effectively, the assumptions allow us to explicitly relate the weighted sum of each premium group's incremental average cost to the "fit" of transfer payments and premium groups to gross costs in the entire population. This is convenient as it allows us to derive a measure of payment system performance with respect to this component of the price distortion without parameterizing a model of consumer demand. The derivation for this expression, along with an expression for ω, is found in Appendix A.

Dividing by ω (since it is the same in all payment system alternatives), we define $1 - \phi = 1 - \frac{\tilde{\phi}}{\omega}$, which we refer to as "payment system fit," the portion of the variance of x that is explained by the revenues allocated to a silver plan by the payment system. We subtract $\frac{\tilde{\phi}}{\omega}$ from 1 merely to make interpretation of the measure similar to the more familiar R-squared measure. All components of this measure can be readily calculated with *ex ante* data on insurance claims and the transfer and premium category rules of a particular payment system.

As now defined and similar to premium fit, as payment system fit increases, the incremental average cost decreases and efficiency increases. If the payment system explains none of the variance in spending (i.e., $\sum_{i=1}^{N}\left[x_i - \left(r_{p,i} + (\bar{x}_a - \bar{r}_{p,a})\right)\right]^2 = \sum_{i=1}^{N}[x_i - \bar{x}]^2$), the gap between gold and silver average costs is unchanged, and the measure is equal to zero. If the payment system explains 100% of the variance in spending at the individual level (i.e., $\sum_{i=1}^{N}\left[x_i - \left(r_{p,i} + (\bar{x}_a - \bar{r}_{p,a})\right)\right]^2 = 0$), reducing the gap between gold and silver average costs to zero, the measure will be equal to one, implying complete elimination of the inefficiency generated by adverse selection.

3.5 Summarizing Inefficiency Due to Inefficient Pricing

We have now derived two measures characterizing the efficiency loss due to inefficient plan pricing. The first, $1 - \delta$, measures the inefficiency due to premium regulations, and the second, $1 - \phi$, represents the inefficiency due to adverse selection. Both measures characterize efficiency relative to a base payment system where there is a single premium. Similar to the conventional R-squared measure of statistical fit, they both range from 0 to 1 with 0 implying no efficiency gain relative to the base payment system and 1 implying complete elimination of the efficiency losses under the base payment system. In principle, these pieces are part of a total welfare loss (see (2)). Summing them to one measure would require knowledge of behavioral parameters regarding demand response to premiums among various cost groups determining the degree of selection, knowledge unavailable *ex ante*.[42]

4. Inefficiencies from Plan Actions

In Section 3, we assumed health plan characteristics were fixed. In practice, insurers structure their products to attract profitable consumers and deter unprofitable ones, creating the second source of selection-related inefficiency in health insurance markets reviewed above in Section 2.3. Plan actions include discriminatory recruitment of consumers with certain chronic illnesses, decisions about what market segments to enter, and loose (tight) rationing of services attractive to low-cost (high-cost) potential enrollees.

In this section, we develop a measure of the relative inefficiency due to insurer incentives to distort service or benefit offerings. Our approach incorporates discrimination based on groups of

[42] If a researcher had estimates of these behavioral parameters, a unified measure of inefficient plan choice could be derived.

people (as in the predictive-ratio strain of this literature) and discrimination based on utilization of groups of services (as in the service-level selection strain). The development of the measure of insurer incentives to distort allocations derives from a condition for plan profit maximization as in earlier papers in this literature, with the primary innovation being a generalization of the basis on which discrimination can take place.

We begin with the smallest units over which a plan might conceivably take actions, and, using these units as building blocks, aggregate as necessary to characterize incentives at the partition of plan health care spending corresponding to actions a plan might more realistically take. We then aggregate the distortionary incentives over all partitions to develop a relative measure of payment system performance with respect to the distortionary incentives the payment system conveys to insurers. We illustrate how the partition-based measure works by applying it to a series of special cases at the close of this section.

4.1 A Partition of Health Care Spending

We regard a health plan as consisting of a set of individual-specific allocations of S medical services. If person i joins the plan, she receives x_{is} of medical service s, measured in dollars. A health plan can thus be described as an $N \times S$ matrix X of individual allocations of medical services:

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1S} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{NS} \end{bmatrix}$$

Total spending on person i for all services is $x_i = \sum_s x_{is}$.

It is clear that $\{1, \dots, N\}$ is the most disaggregated partition of spending possible at the person level and we define $\{1, \dots, S\}$ such that it is the most disaggregated partition of medical spending possible at the service level. Although our building-block partition has $N \times S$ cells, a plan might be able to discriminate only at the service level, or perhaps at the service level for a selected group of potential enrollees. For example, a plan might locate primary care centers to favor enrollees from particular counties. This tactic would correspond to a combination of people (the i's) and services (the s's) in our partition.

Consider a new, coarser partition, Q, grouping people and services according to actions a plan might realistically take, where $Q = \{x^1, \dots, x^Q\}$. We require that each allocation, x_{is}, is in one and only one element of Q, x^q. Figure 2 illustrates the idea of a partition starting with 5 individuals and 5 services, generating 25 allocations of services to individuals. Suppose that insurers can only modify service-level allocations on the inpatient/outpatient margin where services 1 and 2 are inpatient services and services 3, 4, and 5 are outpatient services. Also suppose that insurers can only modify individual-level allocations at the county level where persons 1 and 2 reside in County A and persons 3, 4, and 5 reside in County B, resulting in the partition $Q = \{x^1, x^2, x^3, x^4\}$ where:

$$x^1 = \{x_{11}, x_{12}, x_{21}, x_{22}\}$$
$$x^2 = \{x_{13}, x_{14}, x_{15}, x_{23}, x_{24}, x_{25}\}$$
$$x^3 = \{x_{31}, x_{32}, x_{41}, x_{42}, x_{51}, x_{52}\}$$
$$x^4 = \{x_{33}, x_{34}, x_{35}, x_{43}, x_{44}, x_{45}, x_{53}, x_{54}, x_{55}\}$$

When a plan takes action on one element of x^q, it must take similar action with respect to all other elements of x^q. This implies, for example, that if an insurer desired to increase person 1's allocation of service 1, it must also increase person 1's allocation of service 2 and person 2's allocations of services 1 and 2. In this example, the insurer might perform this action by adding a hospital to its network in County A.

4.2 A Measure of Incentives to Distort Allocations

Let $\text{rev}^j_{p,i}$ be the revenue to a plan associated with person i (in plan j) under payment system p. As above, revenues can be affected by both premiums and transfers. Also as in Section 3, we evaluate payment systems based on the *incentives* they convey to plans to take actions to attract/deter risks, not on the magnitude of the insurer's behavioral response to those incentives, which is unknown *ex ante*. We interpret "actions" as describing any activity directed at deterring/attracting membership from individuals within a partition element. The incentives to take actions are operationalized as resulting from a plan changing spending on an element of a partition.

Efficiency requires an allocation of funds across enrollees and services according to the enrollees' valuation of the service-level allocations, whereas a payment system creates incentives to direct funds towards partitions associated with profitable enrollees and away from partitions associated with unprofitable enrollees. We characterize an efficient allocation of funds across elements of Q, compare this to the allocation implied by profit-maximization based on the incentives created by a payment system, and develop a measure of incentives benchmarked against efficiency that can be operationalized with elements of *ex-ante* individual-level health insurance claims.

N_t is the number of type-t consumers in the population. Type-t consumers have the same level of (expected) spending for each of the S services: $x_{is} = x_{ts}\ \forall i \in t$ Type-t consumers' utility functions over x_{ts} are identical and additively separable across services: $v_t = \sum_s v_{ts}(x_{ts})$.[43] Net social benefits from services are the sum of individual benefits less costs:

$$W = \sum_t N_t \sum_s [v_{ts}(x_{ts}) - x_{ts}]$$

Profits for a plan j, on the other hand, are revenues less costs:

$$\pi^j = \sum_t n_t^j \left[rev_{p,t}^j - \sum_s x_{ts}^j \right]$$

The partition forces a strict relation between spending at the q level and its distribution to the x_{is} level, which we capture by shares σ_{is}^q, which divide each dollar spent on q into spending on (i,s) combinations in a fixed proportion. Hence, $\sum_{is \in x^q} \sigma_{is}^q = 1$ for i,s combinations in cell q and $\sigma_{is}^q = 0$ for all i,s not in cell q. Since health care spending is a private good, we can define $\sigma_{ts}^q = \frac{\sigma_{is}^q}{N_t}$ which divides spending equally among the type-t individuals. Thus, if $\hat{x}^q$ is the total spending on all allocations in cell q, then the spending going to each type-t individual is $x_{ts} = \hat{x}^q \frac{\sigma_{is}^q}{N_t} = \hat{x}^q \sigma_{ts}^q$. Note also that $\sum_{ts} N_t \sigma_{ts}^q = 1$.

Consider the effect of a marginal change of $\hat{x}^q$ on net benefits and profits. In terms of net benefits, a change in $\hat{x}^q$ yields:

$$\frac{\partial W}{\partial \hat{x}^q} = \sum_t N_t \sum_s [\sigma_{ts}^q v'_{ts} - \sigma_{ts}^q]$$

Allocation of a budget across cells is efficient if for any q and $\tilde{q}$ the following equality holds:

$$\sum_t N_t \sum_s \sigma_{ts}^q v'_{ts} = \sum_t N_t \sum_s \sigma_{ts}^{\tilde{q}} v'_{ts} \tag{6}$$

Profit incentives work through consumers' enrollment decisions as a plan seeks to attract and deter enrollees depending on their profitability. Define a type-t individual's utility in health plan j as $v_t^j = \sum_s v_{ts}(x_{ts}^j) + \epsilon_t^j$, where x_{ts}^j is the spending on service s a person of type t gets in plan j, and ϵ_t^j is the individual-specific idiosyncratic element of plan valuation. Let $n_t^j(\sum_s v_{ts}(x_{ts}^j))$ be the number of type-t individuals choosing to enroll in plan j, which is a function of the total allocation of funds to type-t individuals under the plan. For any amount of spending on a q-cell, $\hat{x}^{jq}$, x_{ts}^j follows the partition rule so that $x_{ts}^j = \hat{x}^{jq} \sigma_{ts}^q$.

[43] In Section 3, "type" involved fixing only total expected spending. Here we require types to have the same expected spending for each service, s, and the same utility function for services. There will shortly be heterogeneity within a type with respect to plan valuation.

Using the expression for profits from above, the change in profits with respect to a change in the allocation of funds to cell q is:

$$\frac{\partial \pi^j}{\partial \hat{x}^{jq}} = \sum_t \sum_s \left\{ n_t^{j\prime} v_{ts}' \sigma_{ts}^q \left[rev_{pt}^j - \sum_s x_{ts}^j \right] - N_t^j \sigma_{ts}^q \right\}$$

Assume that plans are in a symmetric equilibrium as in FGM (2000) so the j superscripts can be dropped. Profit maximization implies that an insurer will allocate funds across partitions such that for any q and $\tilde{q}$ the following equality must hold:

$$\sum_t n_t' \sum_s v_{ts}' \sigma_{ts}^q \left[rev_{pt} - \sum_s x_{ts} \right] = \sum_t n_t' \sum_s v_{ts}' \sigma_{ts}^{\tilde{q}} \left[rev_{pt} - \sum_s x_{ts} \right] \tag{7}$$

Recalling our earlier assumption that the enrollment elasticity of demand, $\frac{n_t'}{N_t}$, is constant for all types, after dividing through by the inverse of the constant elasticity, (7) can be rewritten as:

$$\sum_t N_t \sum_s \sigma_{ts}^q v_{ts}' \left[rev_{pt} - \sum_s x_{ts} \right] = \sum_t N_t \sum_s \sigma_{ts}^{\tilde{q}} v_{ts}' \left[rev_{pt} - \sum_s x_{ts} \right] \tag{7$'$}$$

To operationalize this expression we make the assumption that in the *ex-ante* claims data, individual-by-service-level allocations are (second-best) efficient such that $v_{ts}' = v'$ for all t,s.[44] This reduces $(7')$ to the summation of the product of the allocation rule, σ_{ts}^q, and type-t profitability:

$$\sum_t N_t \sum_s \sigma_{ts}^q \left[rev_{pt} - \sum_s x_{ts} \right] = \sum_t N_t \sum_s \sigma_{ts}^{\tilde{q}} \left[rev_{pt} - \sum_s x_{ts} \right] \tag{7$''$}$$

Finally, noting that fund allocations in the data are subject to the partition rule, all components of this expression are found in *ex ante* claims data.

Comparing (6) and $(7'')$ reveals that in this setting the incentive for profit-maximizing firms to distort a q-cell allocation away from the efficient allocation is governed by the correlation between the profitability of types, $\left(Rev_{pt} - \sum_s x_{ts}\right)$, and σ_{ts}^q. For a particular q-cell, as the product of these two terms moves further from zero (in either direction), the equilibrium allocation of funds to the cell diverges from the efficient allocation. Our measure of payment system performance is

[44] Two comments are worth making about this assumption. First, a version of this assumption is commonly, if implicitly, made in the literature that uses *ex-ante* claims data to compare the efficiency properties of health plan payment systems. Use of "predictive ratios" implicitly assumes that the costs for the group in question are the target for payment. See also Frank, Glazer, and McGuire (2000) and papers that follow in the literature on service-level selection. Second, this assumption implies that consumers place equal value on an additional dollar of anticipated service-level spending on *any* service. Consumer differences in the ability to predict spending in different categories could interfere with this assumption. We relax the assumption of equal predictability in the robustness section below.

based on the square of this product, again invoking Harberger-type arguments that the relationship between the distortionary incentive and welfare loss is approximately quadratic.[45]

To characterize the payment system with respect to distortionary incentives over all parts of the Q partition, we construct a sum of the square of the incentives described in Equation (7′′). This generates the following metric:

$$\sum_q \left[\sum_t N_t \sum_s \sigma_{ts}^q \left(\text{rev}_{pt} - \sum_s x_{ts}\right)\right]^2 \tag{8}$$

Similar to Section 3, we normalize our measure to describe the performance of payment system p relative to a simple payment system where $\text{rev}_{p,i} = \bar{x}$ for all i. We also subtract our measure from one to make the interpretation of the metric similar to an R-squared, yielding:

$$\psi = 1 - \frac{\sum_q \left[\sum_t N_t \sum_s \sigma_{ts}^q (\text{rev}_{pt} - \sum_s x_{ts})\right]^2}{\sum_q \left[\sum_t N_t \sum_s \sigma_{ts}^q (\bar{x} - x_i)\right]^2}.$$

Finally, assuming that the fixed proportion by which shares are allocated within a q-cell is governed by the patterns of service use found in the *ex ante* claims data, substituting for $N_t\sigma_{ts}^q$, and defining $\hat{x}_i^q$ as the sum of all of person i's allocations in cell q, $\hat{x}_i^q = \sum_s x_{is} 1[x_{is} \in x^q]$: [46]

$$\psi = 1 - \frac{\sum_q \left(\sum_i \frac{\hat{x}_i^q}{\hat{x}^q} (\text{rev}_{p,i} - x_i)\right)^2}{\sum_q \left(\sum_i \frac{\hat{x}_i^q}{\hat{x}^q} (\bar{x} - x_i)\right)^2}.$$

The $(\text{rev}_{p,i} - x_i)$ term in the numerator is the profit per person. The ratio $\frac{\hat{x}_i^q}{\hat{x}^q}$ captures which individuals are "hit" by an increase in spending in cell q of a partition and therefore encouraged to

[45] We choose the square of the product rather than raising the gap to some other power for reasons similar to those discussed in Section 3. Our social welfare function generates a setting in which the welfare loss due to distorted allocations of funds is proportional to one-half times the square of the difference between the equilibrium and efficient allocations. Assuming a linear relationship between the distortionary incentive and the actual allocation distortion, there should thus be a quadratic relationship between the incentive and welfare.

[46] This relationship can alternatively be derived by considering a setting where insurers set their benefits in competition. First, note that the Harberger-type argument implies that the welfare loss due to distorted allocations of funds is proportional to one-half times the square of the difference between the equilibrium and efficient allocations. Assume that the total spending for an individual in the *ex ante* data is the efficient allocation. Assume that insurers can set allocations at the individual level. Competition will force the insurer to set each individual's equilibrium allocation equal to the revenue attached to that individual. This results in the welfare loss for an individual being proportional to the square of the difference between her revenue and the cost found in the data, a special case of the measure presented here where insurers can select at the individual level. This argument is also generalizable to the more general metric presented here.

join. Incentives work at the q level so the correlations between joining incentives and profits per person must be summed over all q cells to fully characterize incentives. As the payment system does a better job at matching revenues to costs, ψ increases, denoting an improved set of incentives.

4.3 Illustrative Special Cases

We illustrate how our measure works in four special cases of the selection partition, showing how our measure relates to some previously applied measures discussed in Section 2.3. Columns in Table 1 indicate whether the plan can distort allocations at the most disaggregated service level and the rows indicate whether the plan can distort allocations at the most disaggregated individual level. There are four possibilities. The top left combination illustrates the setting where the plan can distort on neither the individual nor the service level and the partition has only one cell which includes all possible allocations:

$$x^1 = X = \begin{bmatrix} x_{11} & \cdots & x_{1S} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{NS} \end{bmatrix}$$

and

$$\hat{x}^1 = \sum_s \sum_i x_{is}.$$

In the bottom right combination the plan can select on both people and services, and the partition has a cell for every possible individual-service combination: $x^1 = x_{11}, x^2 = x_{12}, \dots, x^{N\times S} = x_{NS}$. The upper right cell corresponds to only service-level (but not individual-level) selection and the lower left cell is only individual-level (and not service-level) selection.

There are several notable features of these four special cases. First, the table shows that under *any partition* where insurers can discriminate at the individual-level, the selection measure reduces to the R-squared from a regression of (expected) individual annual medical spending on $rev_{p,i}$.[47] The formula for ψ is the same in both columns for "Yes" individual selection. This is true whether the insurer can discriminate across all services or no services and everything in between those two extremes. This result is due to our assumption that given the allocations in the *ex-ante* claims data, an additional dollar of spending has the same effect on consumer utility, no matter what service that dollar is allocated toward: $v'_{ts} = v'$ for all t,s. Thus, if the insurer can discriminate at the individual-level, it can perfectly manipulate the utility the plan offers the consumer, whether it can distort allocations at the service level or not.

[47] We show in Section 5.3 below that in a setting where premiums are set in competition (and there is no regulated age curve) this is equal to "payment system fit."

Second, the top right cell of the table presents our selection measure in the case where insurers can distort at the service level but not at the individual level, as studied by FGM (2000), Glazer and McGuire (2000), and Ellis and McGuire (2007). Similar to those papers, our measure implies that in this setting the appropriate metric for comparing insurer incentives to inefficiently distort benefits is related to the "predictiveness" of a service, or the correlation between an individual's utilization of the service and her total profitability to the plan.[48]

Finally, in the case where an insurer cannot discriminate at the individual or the service level (the top left cell of the table), there is still potential for efficiency loss, and the loss is related to the correlation between an individual's share of total medical spending in the population ($\frac{x_i}{\bar{x}}$) and her profitability to the insurer ($rev_{p,i} - x_i$). While there is only one cell in the partition, it remains possible for a plan to cut back on all of these services and distort coverage away from the optimal allocation just as originally described by Rothschild and Stiglitz (1976) who modelled coverage versus premium as the mechanism for distortion. In their model, the correlation between demand ($\frac{x_i}{x}$) and cost ($rev_{p,i} - x_i$) results in inefficient contracts being provided in equilibrium (or the lack of an equilibrium altogether). This same correlation drives our measure of inefficiency in this special case.

Moving beyond review of these cases, it is important to note that this measure is customizable for any setting. A researcher or policymaker can choose partitions that align with their beliefs about the margins on which insurers can distort contracts in a particular market. For example, if a policymaker believes insurers can distort benefits only through hospital networks and through strategic prescription drug formulary construction, she may want to partition the data according to drug class and hospital. If she also believes that benefits can be distorted geographically, she may choose to combine the hospital and drug partition with an individual partition that groups people by county. The only requirement is that the partition consists of a mutually exclusive and exhaustive set of cells of the allocations. In the empirical section below, we will provide some examples of potential partitions.

[48] The measure here does not capture the element of predictability also emphasized by FGM (2000) and EM (2007). The implicit assumption is that the degree of predictability does not differ across services. We relax this in the robustness section below.

4.4 Summary of Selection Measures

Table 2 summarizes the three selection-based measures we have derived, capturing both selection-related inefficiency associated with inefficient plan choice, and the inefficiency associated with plan distortion of benefits. The first row relates to the properties of the premium categories themselves and the second row to the combined premium-plan payment system features capturing how well the payment system tracks costs and reduces the gap between gold and silver average (net) costs. These two metrics were derived in Section 3 and they measure welfare loss due to inefficient sorting. The third row presents the general expression for the measure derived in Section 4, the measure of efficiency loss due to insurer incentives to distort plan benefits to attract healthy enrollees. Recall that to operationalize this measure the researcher or policymaker must choose a partition of spending corresponding to her belief about the individual- and service-level margins on which an insurer can discriminate.

The measures vary between 0 and 1 with 0 implying the payment system fails to improve efficiency in relation to a base payment system and 1 implying complete elimination of inefficiency. Notably, the measures have affinities to the measures traditionally used in the literature to capture selection-related inefficiencies under risk adjustment: the R-squared statistic and predictive ratios. The first two measures are basically modified R-squareds. The fit of the premium categories captures one aspect of the sorting problem and the payment system fit is a metric relevant to EF-type selection problems. The third measure is similar to an aggregation of an exhaustive set of predictive ratios that relates distortionary incentives to welfare by squaring the incentive to distort each q-cell allocation.

5. Data and Methods

5.1 Data

To illustrate the computation and interpretation of the selection measures developed in the preceding two sections we put ourselves in the position of a policymaker or regulator designing a payment system for the Health Insurance Marketplaces. We use a more recent version of the health insurance claims data used by Kautter et al (2014) to develop the HHS-HCC Marketplace payment system, the Truven MarketScan Commercial Claims and Encounters dataset (MarketScan) for 2012 and 2013 which includes information on individuals with employer-based group health insurance. The payment system designed for use in the Marketplaces applies separate risk adjustment formulas to children and adults. We focus on adults 21-64, not children. Following criteria applied for estimation of the 2014 HHS-HCC model, we keep individuals enrolled in a preferred provider organization (PPO) or other fee-for-service (FFS) health plan in both the first and last months of a both years,[49] and who have no payments made on a capitated basis (Kautter et al., 2014). Also, following HHS criteria, we require individuals to have both mental health and drug coverage. We exclude individuals who have claims with negative payments in services. Since we will be estimating prospective as well as concurrent risk adjustment models, we require individuals to be enrolled in both 2012 and 2013. After applying the inclusion and exclusion criteria, we have two years of claims for 7,072,964 individuals.

In addition to being more recent and including two years of data, our sample differs from that used in Kautter et al (2014) in another important way. Rather than using the full remaining MarketScan sample, we use a propensity-matched subsample of MarketScan data which yields a sample chosen to more closely reflect the age, gender, geography and disease characteristics of people eligible for coverage in the ACA Marketplaces. Adults covered by group health insurance in MarketScan will, for example, likely be older and sicker than those in the Marketplaces. We anticipate that the selected sample, with over 2 million people, will improve calibration of risk adjustment models and yield more valid simulations of likely Marketplace incentives by using data on individuals more similar to Marketplace enrollees.[50]

We select this subset from MarketScan using methods from earlier work using the Medical Expenditure Panel Survey (MEPS) to define persons with the characteristics that would make them eligible for the Marketplace (McGuire et al., 2013, 2014), and propensity model methods within MEPS using variables common to MEPS and MarketScan (Rose et al., forthcoming). In essence we estimate the propensity an observation in MEPS is Marketplace eligible (in relation to large group health insurance) and then use this propensity score, along with nationally representative population weights from MEPS to draw a sample from MarketScan, which we refer to as the MarketScan Marketplace (MM) sample. Among the advantages of this sample is that it is nationally representative and can be used for estimation and simulation without further weighting. Appendix C to this paper has more details on our matching and selection methods.

---

[49] Other FFS plans include Exclusive Provider Organization, Non-Capitated Point-of-Services, Consumer-Driven Health Plan, and High-Deductible Health Plan.

[50] At some point, claims and enrollment data will be available for at least some Marketplaces. It is unlikely that such a large sample will be available in time for any recalibration of payment models for 2017.

Summary statistics for our MM sample and, for purposes of comparison, a random sample from the full MarketScan sample (meeting our inclusion/exclusion criteria and from which our sample was drawn) are shown in Table 3. Most of the variables are self-explanatory. Among the variables we used in the propensity score are indicator variables for the presence of any of ten chronic illnesses from any time during the two years of observation. As anticipated, the full MarketScan sample was older and had a higher prevalence of all chronic conditions than the MM sample. Observations in our MM sample spent less on average than the full MarketScan data though this was not true for each major spending category.

5.2 Payment Systems

We calculate our measures under payment systems that differ in three dimensions: risk adjustment, reinsurance, and premium regulation. The payment systems were chosen because they are of potential policy interest and also to illustrate important features of the selection measures. The payment systems are listed in Table 4.

The first column of Table 4 is ACA 2017, the proposed payment system for Marketplaces in 2017 and beyond. For this case, we assume that Marketplaces continue to apply the federally recommended age curve for premiums, risk adjustment is concurrent based on the specification of the HHS-HCC model, and the compulsory reinsurance system in place for 2014-2016 is discontinued as was stipulated in the ACA. We also apply the risk adjustment "transfer formula" used by HHS to coordinate risk adjustment and premiums when both include age adjustments.

The next three columns in Table 4 describe three alternative payment systems. Alternative 1 changes risk adjustment to be prospective (based on prior year diagnoses) rather than concurrent (based on current year diagnoses) using the same HHS-HCC specification. Prospective diagnostic risk adjustment was not feasible in 2014, but it is for 2017, as diagnoses from prior year claims are now available. We implement prospective risk adjustment combined with simple age-gender risk adjustment for new enrollees, mimicking the approach used by Medicare to pay Medicare Advantage plans for new enrollees. We assume a random 50 percent of all Marketplace enrollees to be new enrollees, and use the age-gender only model for them. As with the ACA 2017 payment system, we maintain the HHS transfer formula to avoid over-adjusting for age.

Alternative 2 is the same as Alternative 1 but restores the reinsurance program in force in the Marketplaces during 2014, paying 80% of costs over an attachment point of $60,000 in plan costs.[51] Finally, Alternative 3 is similar to Alternative 2 in that it uses prospective risk adjustment (with an age-sex model for new enrollees) combined with reinsurance, but allows premiums for all ages to be set in market equilibrium rather than according to a pre-specified age curve.

5.3 Estimation of Selection Measures

The selection measures call for individual level data on x_i, r_{pi}, and rev_{pi} under each payment system. For now, we assume that consumers and plans have perfect foresight so that the expected cost, x_i, is equal to *ex post* medical spending, which is observed in the data. We will relax this assumption later and consider the sensitivity of our findings to assumptions about cost predictability. r_{pi} and rev_{pi} must be simulated under each payment system, taking into account risk adjustment, reinsurance, and the risk adjustment transfer formula. We explain how we simulate each of these components of r_{pi} and rev_{pi} and outline the partitions of medical spending we choose for estimating ψ from Section 4.

Risk Adjustment

Under all payment systems we consider, an individual's risk score is defined (following HHS) as a linear combination of a vector of individual risk adjustment variables and a vector of weights assigned to those variables: $risk_{pi} = \beta z_i$. In the ACA 2017 payment system, z_i consists of the same set of variables used in the HHS-HCC risk adjustment model used in the Marketplaces: a set of dummy variables indicating the individual's age-by-sex cell and any recognized conditions she received a diagnosis for during the current year. Three risk adjustment models are calibrated: a concurrent model that mimics the estimation used in the existing HHS-HCC concurrent risk adjustment model that consists of indicators for HCCs (along with age and sex) from 2013 to predict 2013 spending, a prospective model with the same structure but which uses HCCs from 2012 to predict 2013 spending, and an age-sex only model that is used for predicting spending for "new enrollees" without information on prior diagnoses.[52]

---

[51] Some form of this reinsurance policy is in place in the Marketplaces for the first three years of operation (2014-16), though the parameters of the policy vary for each year, with the generosity of the policy decreasing monotonically.

[52] While the actual weights used in the Marketplaces are publicly available, we estimate our own weights so that the comparison of concurrent and prospective risk adjustment are "apples-to-apples."

We apply HHS software to the data to generate HHS-HCCs for 2013 and 2012 for estimating the concurrent and prospective models, respectively.[53] For the simulated payment systems, these HCCs along with a set of age-by-sex cells also included in the HHS-HCC risk adjustment model make up the vector of risk adjustment variables, z_i. The vector of weights, β, is estimated for each model via a regression of total 2013 spending on z_i for the full MM sample. To avoid "overfitting," we use a k-fold estimation procedure by which the weights used for calculating a particular individual's risk score are estimated using data only on *other* individuals.[54] We also estimate an age-gender only model on data from 2013, with the same age and gender categories used in the HHS-HCC risk adjustment formula. All three models are normalized to have a mean risk score of 1.0 in the MM sample.

To simulate plan payments with prospective risk adjustment, we recognize that a share of the population will not have prior year diagnostic data available by designating a random share (we assume 50%) of the population to be "new enrollees." We call the others "incumbent enrollees" for whom we observe claims from the prior year. For the new enrollees, z_i consists of only the age-by-sex cells and not the chronic condition indicators. For the incumbent enrollees, z_i consists of the same set of dummy variables as in the ACA 2017 payment system case, but the chronic condition indicators are based on diagnoses from the prior year rather than the current year. The β-weights are also estimated separately for the prospective model used for the incumbent enrollees and the age-gender model used for the new enrollees. Estimated coefficients for the three risk adjustment models using the MM sample are reported in Appendix D.

Transfer Policy

The Marketplace transfer policy is intended to take account of the age structure of premiums in risk adjustment transfers (Pope et al. 2014). Consistent with this policy, for the ACA 2017 payment system and Alternatives (1) and (2), we define the risk adjustment transfer, $\mathrm{ra}_{\mathrm{pi}}$, as follows:

$$\mathrm{ra}_{\mathrm{pi}} = \left(\frac{\mathrm{risk}_{\mathrm{pi}}}{\overline{\mathrm{risk}}_{\mathrm{p}}} - \frac{\alpha_{\mathrm{i}} - \overline{\alpha}}{\overline{\alpha}}\right)\overline{\mathrm{x}}.$$

In the expression, $\mathrm{risk}_{\mathrm{pi}}$ is person i's risk score and $\overline{\mathrm{risk}}_{\mathrm{p}}$ is the population average risk score under payment system p. Similarly, α_{i} is person i's "age weight" and $\overline{\alpha}$ is the population average age

[53] Software can be found here: http://www.cms.gov/CCIIO/Resources/Regulations-and-Guidance/Downloads/SASsoftware.zip. For simplicity, we do not impose constraints on the coefficients as was done by Kautter et al (2014) to ensure monotonicity within each hierarchy and nonnegative coefficients.
[54] For all models, we set k=10. The k-fold procedure is described in more detail in Appendix C.

weight. Under these payment systems, the age weights are defined by the age curve specified by HHS. The curve is presented in Table D2 in Appendix D with values ranging from 1.0 for 21 year-olds to 3.0 for 64 year-olds. Under Alternative (3), the zero-profit market equilibrium makes an age-adjustment unnecessary in a transfer policy.

Reinsurance

Consistent with the reinsurance policy in place in the Marketplaces during 2014, we define reinsurance payments, repay_{pi}, to be 80% of an individual's annual spending above \$60,000. While the Marketplace policy is not required to be budget neutral, we enforce that here and require that the average value of the net reinsurance payment, re_{pi}, be equal to zero. To implement this, we calculate the insurer's constant actuarially fair reinsurance premium per person, reprem, define the net reinsurance payment as $\mathrm{re}_{pi} = \mathrm{repay}_{pi} - \mathrm{reprem}$, and include this adjustment in the net revenue calculation.

Revenues and Transfers

Measuring premium fit and payment system fit requires x_i, r_{pi}, and rev_{pi}. As discussed above, for now we define x_i as *ex post* spending as it appears in the claims data. We define r_{pi} to account for all transfers and the transfer formula itself. First, denote the concurrent normalized (relative) risk score as $\mathrm{RRS}_{pi}^{conc} = \frac{\mathrm{risk}_{pi}}{\overline{\mathrm{risk}}_p} = \frac{\beta^{conc} z_i}{\beta^{conc} \bar{z}}$. Then, for each payment system:

ACA 2017: $r_{pi} = \left(\mathrm{RRS}_{pi}^{conc} - \frac{\alpha_i - \bar{\alpha}}{\bar{\alpha}}\right)\bar{x}$

Alternative (1): $r_{pi} = \left(\mathrm{RRS}_{pi}^{pros} - \frac{\alpha_i - \bar{\alpha}}{\bar{\alpha}}\right)\bar{x}$

Alternative (2): $r_{pi} = \left(\mathrm{RRS}_{pi}^{pros} - \frac{\alpha_i - \bar{\alpha}}{\bar{\alpha}}\right)\bar{x} + \mathrm{repay}_{pi} - \mathrm{reprem}$

Alternative (3): $r_{pi} = \mathrm{RRS}_{pi}^{pros}\bar{x} + \mathrm{repay}_{pi} - \mathrm{reprem}$

For the measure developed in Section 4, ψ, we also need to define rev_{pi}. The risk adjustment and reinsurance transfers defined above combine with premiums to generate the total revenue a plan receives for enrolling an individual. Revenues can be expressed as $\mathrm{rev}_{pi} = \mathrm{trans}_{pi} + \mathrm{prem}_{zi}^{j}$. Recall that in Section 4 we assumed that insurance plans were in a symmetric equilibrium. This implies that all plans have the same premium. Under payment systems that use the regulated age curve, the premium for individual i will be equal to a base premium multiplied times the age weight, $\mathrm{prem}_i = \widehat{\mathrm{prem}}\alpha_i$. We assume that competition forces premiums to be equal to average cost on average, producing the result that the premium for individual i can be written as $\mathrm{prem}_i = \bar{x}\frac{\alpha_i}{\bar{\alpha}}$. Transfers are defined by the risk adjustment and reinsurance policies above so that for a payment system incorporating both risk adjustment and reinsurance, $\text{trans}_{\text{pi}} = \text{ra}_{\text{pi}} + \text{repay}_{\text{pi}} - \text{reprem} - \bar{\text{x}}$. The exact expressions we use to construct revenues are found by the appropriate substitutions and are written out here. These correspond to the four simulations listed in Table 4:[55]

ACA 2017: $\text{rev}_{\text{pi}} = \text{RRS}_{\text{pi}}^{\text{conc}}\bar{\text{x}}$

Alternative (1): $\text{rev}_{\text{pi}} = \text{RRS}_{\text{pi}}^{\text{pros}}\bar{\text{x}}$

Alternative (2): $\text{rev}_{\text{pi}} = \text{RRS}_{\text{pi}}^{\text{pros}}\bar{\text{x}} + \text{repay}_{\text{pi}} - \text{reprem}$

Alternative (3): $\text{rev}_{\text{pi}} = \text{RRS}_{\text{pi}}^{\text{pros}}\bar{\text{x}} + \text{repay}_{\text{pi}} - \text{reprem} + \bar{\text{x}}_{\text{a}} - \left(\overline{\text{RRS}}_{\text{p}}^{\text{pros}}\bar{\text{x}} + \overline{\text{repay}}_{\text{pa}} - \text{reprem}\right)$

All of the components of rev_{pi} for each payment system are found in our data, with the exception of $\text{RRS}_{\text{pi}}^{\text{conc}}$ and $\text{RRS}_{\text{pi}}^{\text{pros}}$ which we estimate through regressions as discussed above. Note that the transfer formula is used in the definition of rev_{pi}, but that the age adjustments in the formula and the age adjustments to premiums cancel each other out.

Partitions for Benefit Distortion Measures

We choose three partitions of spending that generate different variations on the general measure of insurer incentives to distort plan benefits. We denote these three variatons ψ_1, ψ_2, and ψ_3. For ψ_1, we assume that insurers can discriminate against individuals with four common chronic diseases (cancer, diabetes, heart disease and mental health) separately, generating a partition with 5 cells, one for each chronic disease containing all of the spending among individuals with that disease and one containing all spending for individuals with none of the chronic diseases. We classify individuals into chronic disease groups using diagnoses. If a person has a claim for more than one chronic illness, we place that person in the category with the largest amount of spending so that the disease groups are mutually exclusive.

For ψ_2, we assume that insurers can manipulate medical services at the level of three broad categories: inpatient, outpatient, and prescription drug services. This categorization is consistent with the insurer's ability to set different levels of cost sharing for these broad service categories. Notably, for ψ_2 we only allow insurers to manipulate benefits at the service-level, not at the person-level. For ψ_3, we maintain the insurer's ability to manipulate benefits at the service-level but now allow insurers to manipulate service-level benefits separately for the old and the young, doubling the

[55] Note that for the ACA 2017 payment system and Alternatives (1) and (2) the age weights from the risk adjustment transfer formula and the age weights from the premium expression cancel each other out, implying that in those cases the variation in revenues is determined entirely by the risk adjustment and reinsurance payments. The age-based premium variation will matter only in the premium fit measure.

number of cells in the partition. For each of the four payment systems we calculate ψ, the measure of incentives to inefficiently distort benefits three times, once for each partition.

6. Results

This section presents estimates of our three selection measures and compares them to the commonly applied R-squared and predictive ratio measures. We also present results regarding the robustness of the estimated measures to various assumptions about the specification of *ex ante* expected costs, x_i. Bootstrapped confidence intervals are presented for all measures.

6.1 Basic Results

Figure 3 displays our first two measures of efficiency for the ACA 2017 payment system and three alternatives. A value of 0 on a measure implies no improvement over the base payment system, while a value of 1 implies that the given payment system completely eliminates the inefficiency corresponding to that measure. Error bars show bootstrapped 90% confidence intervals.

The dark bars represent premium fit, which captures the inefficiency due to premium regulation, i.e., the portion of the variance in costs that is explained by the second-best premiums given the set of premium regulations embedded in the payment system (here, this is equivalent to the portion of the variance explained by the weights of the age-curve, as derived in Appendix B). The low level of the dark bars indicate that none of the payment systems considered do much in the way of matching premium categories to costs. The ACA 2017 payment system and the first two alternatives all have identical premium fit because they all rely on the same government-issued age curve. Under all of these payment systems, an insurer sets the price that a 21 year-old would pay, and then the premiums for all ages are the prescribed multiples of that age 21 bid.[56] Because the reinsurance and risk adjustment payments modeled here are all budget neutral, the entire age structure of premiums is the same under these three payment systems. For the ACA 2017 and first two alternative payment systems, the premium fit measure reflects the portion of the variance in costs explained by the age curve, which is precisely estimated to be 0.013.

For the third alternative payment system – which encompasses prospective risk adjustment plus reinsurance and age-based premiums freely set by insurers – the point estimate for premium fit is slightly higher than for the other payment systems, implying greater efficiency. This result is due

[56] The age curve was estimated by HHS and requires 64 year-olds to pay 3 times the premium of 21 year-olds.

to the fact that the federal age curve does not perfectly match the true age gradient in the claims data. However, the difference in premium fit is small, and the 90% confidence interval includes the point estimates of premium fit for ACA 2017 and Alternatives (1) and (2). This implies that a much more radical reform of premium setting than freeing up the age curve would be necessary to make much of an improvement in premium fit.

The light bars are payment system fit, where there is much more action across the alternatives. For the ACA 2017 payment system, this measure is simply equal to the conventional R-squared statistic from a regression of costs on a modified concurrent risk adjustment risk score where the age weights have been removed (because of the way the transfer system works as explained in Section 5.3). We estimate this measure to be equal to 0.432, consistent with, though larger than, the 0.350 R-squared reported by Kautter et al. (2014).[57] Payment system fit for prospective risk adjustment (Alternative 1) is much smaller, 0.065, implying lower efficiency. This is lower than the typical R-squared for a prospective risk adjustment system reflecting our assumption of 50% turnover in Marketplace enrollment and the use of age/gender risk scores for new enrollees. Adding reinsurance to prospective risk adjustment (Alternatives (2) and (3)), moves payment system fit much higher: 0.73.[58] The estimates of both premium fit and payment system fit are also found in Table 5.

Figure 4 presents the measures developed in Section 4 that capture insurer incentives to distort benefits to attract healthy enrollees and avoid sick ones. These measures also range between zero (no inefficiency reduction) and one (eliminating inefficiency). As with payment system fit, the results imply that the ACA 2017 payment system outperforms prospective risk adjustment (Alternative 1). However, prospective risk adjustment plus reinsurance (Alternative 2) again dominates the ACA 2017 payment system. Alternative (3), which allows insurers to freely price by age, not surprisingly performs almost identically to Alternative (2): the improved payment system fit through premium flexibility does not affect benefit or service-level selection incentives because plan

---

[57] We expect our payment system fit to differ from that in Kautter et al. (2014) because our sample reflects the different age, illness and spending distribution expected to better match the Marketplace sample and we predict plan payments without adjusting for levels of cost-sharing (as in gold, silver, etc.) plans as was done in Kautter et al. (2014). We also do not tamper with the estimated coefficients. Finally, our payment system fit is based on modified risk scores where the age weights have been removed, while the R-squared in Kautter et al. (2014) is based on unmodified risk scores.

[58] While the point estimate for Alternative (3) exceeds the estimate for Alternative (2), the confidence intervals for these measures overlap, suggesting the actual performance of these two payment systems with respect to this type of distortion may not differ.

revenues for a given individual are effectively unchanged between Alternatives (2) and (3). In sum, across our four alternative partitions, prospective risk adjustment combined with reinsurance (Alternative 2) dominates the other payment systems on all measures, supporting the view that reinsurance plus prospective risk adjustment improves upon the concurrent risk adjustment model alone, even when the claims necessary to generate diagnosis-based prospective risk scores are only available for 50% of the enrollees. The estimates of these measures are also presented in Table 5.

6.2 Comparison to Conventional Metrics

It is informative to compare our new metrics with the conventional metrics used to evaluate payment systems: the R-squared for the risk adjustment model and predictive ratios for selected disease categories. The R-squared measures are presented in Figure 5. The ordinary R-squareds from concurrent and prospective risk adjustment are slightly larger than our estimates of payment system fit for ACA 2017 and Alternative (1) (prospective risk adjustment only), respectively. In these settings the R-squared and payment system fit vary in two ways. First, while the R-squared is based on raw risk scores, payment system fit takes into account the transfer formula by removing the age-weights from the risk scores as discussed in Section 5.3. Second, as discussed in Appendix B, in these settings payment system fit accounts for the fact that the regulated age curve effectively magnifies the EF price distortion for the old and shrinks it for the young by including a scaling factor equal to $\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\alpha_i}{\bar{\alpha}}\right)^2$. As the scaling factor is likely to be close to 1, the removal of the age weights is more important than the inclusion of the scaling factor, and it is easy to see how the removal of the age-weights would result in payment system fit being less than the R-squared.

After adding reinsurance in Alternatives (2) and (3) the difference between the R-squared and payment system fit grows substantially. This is due to the fact that in addition to ignoring the consequences of the regulated age curve, the conventional R-squared differs from payment system fit by ignoring the contribution of reinsurance in matching revenues to costs. The conventional R-squared measure can be seriously misleading when a payment system incorporates features other than regression-based risk adjustment. The R-squared measures are also found in Table 6.

Figure 6 presents the predictive ratios for four disease categories commonly used to assess insurer incentives to discriminate against individuals with chronic conditions: mental health, cancer, heart disease and diabetes. These ratios are closest in spirit to our ψ_1 measure, where the plan was assumed to be able to discriminate according to spending on these four disease areas. A ratio close to 1 implies that payment for the group is equal to the group's cost. A ratio below (above) one implies under(over)payment. Conventional predictive ratios are constructed by dividing the average risk adjustment transfers for the group by the average cost for the group, as is done for the predictive ratios reported in Figure 6 for the ACA 2017 payment system and prospective risk adjustment (Alternative (1)). For Alternatives (2) and (3), we report "payment system predictive ratios" that differ from the conventional predictive ratios in that the payment system ratios account for reinsurance. We construct these payment system predictive ratios by dividing the average revenue an insurer receives for enrolling a member of a group (rev_{pi}) by the average cost of the members of the group. For Alternatives (2) and (3), the unmodified conventional predictive ratios would be identical to those reported for Alternative (1) because all three payment systems incorporate the same risk adjustment model.

Consistent with the findings for the Q1 partition reported above, all conventional predictive ratios are better under the ACA 2017 payment system than under prospective risk adjustment alone (Alternative (1)). Additionally, the "payment system predictive ratios" improve with the addition of reinsurance (Alternative (2)). However, the payment system predictive ratio analysis diverges form our results in Figure 4/Table 5 comparing the ACA 2017 payment system to prospective risk adjustment with reinsurance (Alternative (2)). The chronic disease payment system predictive ratios for prospective plus reinsurance range from 0.70 to 0.82, which are further from one than the predictive ratios for the concurrent, ACA 2017 payment system, which range from 0.81 to 0.96, implying lower suggested efficiency. All predictive ratios are also found in Table 6.

There are at least two potential sources of these divergent results. First, the payment system predictive ratios include everyone with any positive spending in a group (e.g., those with mental health problems) and treat low and high-spenders with equal importance when computing the predictive ratio. Our measure weights each person by spending in the category in question (based on the greater importance of the spending category for persons who spend more). Second, the payment system predictive ratio consists of revenue for the numerator and costs for the denominator. In contrast, by squaring this measure for each disease group to go from measuring incentives to measuring welfare, our measure penalizes larger group-level discrepancies between revenues and costs more than small ones. Reinsurance directs revenues to the highest-cost people and gets more credit in a metric based on squared deviations which are much larger among the higher costs groups.

6.3 Robustness to Assumptions about Expectations of Cost

Up to this point, our estimates of the selection efficiency measures make no distinction between realized and expected spending. Note that both the conventional measures of R-squared and predictive ratios also ignore the distinction between actual versus expected costs. Many have argued that because much of spending is not likely to be predictable, the fact that these conventional measures ignore this issue is a critical shortcoming. However, without a model of the economic behavior that generates inefficiencies, it is unclear (1) why predictability of spending is important and (2) from whose point of view (consumers, firms, or society's) predictability should be considered. In this section, we first discuss where our economic framework suggests predictability matters and then test the robustness of the set of our measures for which predictability is relevant to alternative specifications of expected costs. A primary objective is to check if the results favoring reinsurance withstand this robustness check. Reinsurance payments are directed to the high-end portion of the spending distribution, and if this spending is less predictable, our results favoring reinsurance may need reconsideration.

Deriving measures based on economic behavior guides analysis of where expectations matter and whose expectations are relevant for each of our measures. With respect to premium fit, recall that we showed that an individual's first-best premium is equal to the expected cost for her type. This expected cost was the "rational expectation," or the expectation from the social planner's point of view, and incorporates all information available to society *ex ante*. The reason for this was that in Section 3 we focused solely on a "total surplus" concept of welfare based on the difference between the individual's valuation of the plan and the cost to society of providing it. The same logic implies that the second-best prices to which the first-best prices are compared should also be based on the expected cost from the social planner's point of view. This implies that premium fit should be calculated using expected costs from the social planner's point of view.

The expected costs incorporated into payment system fit are different. Recall that this component captures the difference between the equilibrium price and the second-best price. This implies that this component is affected by the mechanism by which prices are set, which involves market forces rather than the specification of expected costs.[59] Importantly, the only assumption in

[59] We should note that the second-best price is still defined according to the social planner's expectations. However, for most relevant payment systems premium groups are large, resulting in approximately the same second-best prices (gamma times the average cost in the premium group) under any specification of the social planner's expectations (including perfect foresight, i.e. *ex ante* expected costs = *ex post* realized costs) because the non-predictable component of spending will average out.. This is especially true in the payment systems

the derivation of this measure related to expectations was that *in equilibrium* competitive market forces result in insurers setting prices equal to the average realized cost of those consumers choosing the gold plan and the average realized cost of those consumers choosing the silver plan. This assumption is reasonable and is consistent with the assumptions made by EFC (2010). Consumer demand behavior, also influenced by expectations, comes in through the *response* to equilibrium premiums. Consumer response and expectations have a scaling effect on the welfare loss due to EF-type inefficiency, but in our relative measures, behavioral response cancels out. Thus, as long as insurers get averages right in equilibrium, different assumptions about expectations are not relevant for the payment system fit measure representing the EF component of selection-related inefficiencies.[60] Instead, this measure can be based on *ex post* realized costs as in Section 6.1.

Finally, consider ψ, the measure derived in Section 4 related to incentives for plans to distort service offerings across people or services. Recall that this measure was derived from a model of plan profit maximization, where plans set allocations of different services to different individuals based on the effect of altering an allocation on plan profits, and that effect depends on the correlation between an individual's willingness-to-pay for a particular set of services and her profitability (For example, if consumers were unable to forecast their likely use of a particular set of services, there would be no value to a plan in attempting to use those services for purposes of selection.). Plans' expectations matter here, but note that plans must anticipate what consumers expect in order to make decisions about service offerings to maximize profits. This implies that for this measure, consumer expectations, and the plan's expectation of those consumer expectations, are relevant.

It is unknown both how "predictable" spending is from the social planner's point of view (relevant for premium fit) and how well consumers can forecast their spending (relevant for ψ), so we test the robustness of our two measures that incorporate expectations to various specifications of expected spending. For ψ, we make no distinction between consumers' expectations and plans'

---

we simulate here where either everyone is effectively in the same premium group or premiums are allowed to vary by age.

60 This argument is similar to the insight that in the EF framework the incremental average cost curve is based on insurer expectations of the spending of individuals enrolling in each plan given a price P. Different payment systems will shift the incremental average cost curve, but the demand and incremental marginal cost curves will remain fixed across payment systems.

beliefs about those expectations.[61] We also do not separately specify consumers' expectations and the social planner's expectations. Instead, for the measures affected in any way by expectations, premium fit and ψ, we will consider the effect of a range of assumptions from perfect foresight to being able to predict only on the basis of age, gender, and prior utilization. Our measure of expected spending is generated by a linear regression predicting total medical spending in 2013 using the age-by-sex cells from the risk adjustment models and indicators for each consumer's decile of 2012 inpatient, outpatient, and prescription drug spending.[62] We report estimates of our measures where we specify expected spending as a weighted average of our "predicted" spending and *ex post* medical spending, with weights equal to 0%, 20%, 40%, 60%, 80%, and 100%. The results from the sensitivity analysis are found in Figure 3. Each panel of the figure shows a different measure. Estimates for the selection measures for each payment system are on the y-axis and the weight put on *ex post* spending relative to predicted spending is on the x-axis. The top-left panel presents the results for premium fit. ACA 2017 and Alternatives (1) and (2) always have identical estimates for this measure, shown by the solid black line. Alternative (3) always has slightly lower values, indicating higher efficiency, though the gap is decreasing in the weight on *ex post* spending, implying that the federal age curve matches the *ex post* age gradient better than the age gradient based on predicted spending. This is not surprising given that our specification for predicted spending includes only 5-year age bins as predictors, while the age curve is based on 1-year age bins. The other three panels present the results for the Section 4 measures, ψ, with the top-right panel presenting results for the ψ_1 measure (chronic disease groups), the bottom-left presenting the ψ_2 measure (service-level discrimination), and the bottom-right presenting the ψ_3 measure (service-by-age-level discrimination). The results are similar for all three versions of ψ: Alternative (1) is always dominated by ACA 2017 and alternatives with reinsurance dominate ACA 2017 for all of the relative weighting specifications except for the specification where expected spending is equal to predicted spending and the weight on *ex post* spending is 0. This implies that the results reported above regarding the

[61] We have no empirical basis for making any distinction. This is the approach also used in FGM (2000) and related papers. Additionally, it seems reasonable to assume that over time insurers learn (the relevant dimensions) of consumer expectations so that in equilibrium they know consumer expectations precisely.
[62] Instead of traditional deciles, we form one cell including all individuals with $0 of spending and then 9 equally sized cells of individuals with positive spending. Details from the prediction regressions can be found in Appendix E. Note that for ψ_2 and ψ_3 we will need to construct separate measures of expected inpatient, outpatient, and prescription drug spending in addition to the measure of expected total spending. To do this, we use the same prediction regressions with 2013 total service-level spending on the left-hand side rather than 2013 total spending.

ACA 2017 payment system being dominated by prospective payment with reinsurance are robust to reasonable assumptions about the social planner's and consumers' ability to predict future medical spending. Only in the case where no part of spending beyond what is predictable from demographics and deciles of prior spending is predictable by the social planner or by consumers does the ACA 2017 payment system dominate Alternatives (2) and (3). We interpret this result as validation that reinsurance plus prospective risk adjustment is likely to outperform concurrent risk adjustment alone with respect to price and service-level distortions caused by adverse selection. This result also implies that across a fairly wide range of payment system alternatives, our measures will produce similar results when estimated using either simple *ex post* realized costs or using more complicated specifications of expected costs. This result may be important for policymakers designing health plan payment systems under constraints that may prevent the use of complex methods for producing estimates of expected costs.

7. Discussion

This paper develops metrics of incentives for selection that we argue are valid, complete and practical to use for the *ex ante* evaluation of the efficiency of alternative payment systems. By *valid* we mean our measures are derived from economic principles rather than based on *ad hoc* statistical criteria, and parallel the Harberger-type measures of inefficiency driven by the square of the gap between equilibrium and efficient prices. The measures reflect the primary sources of inefficiency caused by adverse selection in individual health insurance markets. By *complete* we mean these metrics allow for comparisons among payment systems that differ on a wide variety of margins including premium regulation, risk adjustment, mixed payment systems, or reinsurance. By *practical* we mean our measures can be readily calculated from the data typically available for an *ex ante* assessment of plan payment alternatives. We have conducted such an assessment here on an updated (and better-selected) version of the data likely to be used for any recalibration of Marketplace plan payments for 2017, demonstrating the practicality of the metrics.

Our first two measures (premium fit and payment system fit) capture inefficiencies due to price distortions caused by premium regulation and adverse selection. These measures are based on how well premiums and total plan revenues match costs at the individual level. Our third measure captures inefficiencies caused by insurer incentives to distort benefits to attract profitable enrollees. This measure requires that the researcher or policymaker specify the dimensions of spending for which distortion by insurers is feasible. The measure is then based on the correlation between individuals' distort-able spending and their profitability to the insurer, also known as the "predictiveness" of the spending dimension. Higher correlations between spending on distort-able dimensions of spending and profitability imply stronger incentives to distort plan benefits and larger potential inefficiencies.

We demonstrate the usefulness of our measures by comparing three alternative payment systems to the payment system slated for use in the ACA Marketplaces in 2017. Our analysis shows that while a prospective risk adjustment model performs notably worse than the proposed concurrent model on all selection metrics, this deficiency is more than overcome when prospective risk adjustment is combined with reinsurance similar to that in use in the Marketplaces during 2014. This result regarding the value of reinsurance for combatting distortionary incentives caused by adverse selection is consistent with other recent research bearing on reinsurance in the Marketplace context (Zhu et al. 2013, Geruso and McGuire 2015). This result about reinsurance is robust to assumptions regarding the portion of the variance of medical spending that is "predictable" by consumers.

A comparison of our metrics of relative payment system performance to more conventional metrics typically used in the literature highlights differences between the measures and identifies where the conventional measures may be misleading. Unsurprisingly, when evaluating payment systems using measures of the statistical fit of a risk adjustment model, the benefits of other payment system features are ignored. More interestingly, we find that when evaluating payment systems using predictive ratios for important disease groups, the benefits of reinsurance are undervalued, even when accounting for reinsurance payments in the ratios. This discrepancy can be explained by two differences between predictive ratios and the welfare-related measures we derive from a model of plan profit maximization. First, while both measures are based on the profitability of attracting a particular group, predictive ratios effectively weight the profitability of each member of the group equally. We show that plan profit maximization implies that members of a group should be weighted by their level of spending relative to the total spending of the group. This weighting is captured by our measure but not by predictive ratios, explaining part of the undervaluation of reinsurance by the predictive ratio measure of incentives. Second, predictive ratios evaluate welfare loss due to an insurer's incentive to discriminate against groups of individuals using raw averages of spending and simulated revenues at the group level. However, basic economics implies that welfare loss is proportional to the square of the distortion and thus also proportional to the square of these measures of relative profitability of the different groups, causing the profitability of the most expensive enrollees to be weighted even more heavily and driving a larger wedge between the value of reinsurance implied by predictive ratios and the value implied by our model-based measure of welfare loss due to insurer incentives to discriminate against groups of enrollees.

Our approach is subject to a number of limitations. The most important is that our measures do not evaluate welfare losses in standard units such as dollars. This implies that the measures cannot easily be used to explicitly model trade-offs among the multiple inefficiencies we focus on (price and benefit distortions due to adverse selection) or between these inefficiencies and other objectives of the payment system. For example, imposing a steeper federal age curve, say 5-1, rather than the current 3-1 maximum between 64 and 21 year olds, will likely better address adverse selection problems and improve welfare according to our measures. It may, however, weaken the perceived fairness of the payment system by requiring older (generally sicker) enrollees to pay much more than the young and healthy. This lack of a standard unit unfortunately prevents us from using these measures to find an "optimal" payment system that maximizes efficiency over multiple objectives. It is worth keeping in mind that this limitation is shared with the conventional *ex ante* measures of insurer incentives we seek to improve upon such as the conventional R-squared and predictive ratios.

Other measures of payment system performance can be considered along with those we develop here, even if they cannot be added together in the same units. For example, our measures can be combined with measures of insurer moral hazard to determine whether a particular payment system dominates other systems. Concerning insurer moral hazard, Geruso and McGuire (2015) measure the "power" of a payment system, capturing incentives for cost control. Results of the power analysis show that a payment system incorporating prospective risk adjustment along with reinsurance dominates a payment system consisting only of concurrent risk adjustment with respect to payment system fit and power. Our results complement their finding by showing that prospective risk adjustment plus reinsurance also dominates concurrent risk adjustment alone with respect to premium fit and our measure of welfare loss due to insurer incentives to distort plan benefits.

In addition to objectives related to fairness and insurer moral hazard, our measures and analysis also abstract from concerns about the "game-ability" of payment system features such as risk adjustment (Kronick and Welch 2014, Geruso and Layton 2015) and from protection against "reclassification risk" (Handel, Hendel and Whinston 2015). While these issues are important and should also be taken into account when evaluating payment systems, they are beyond the scope of this paper.

Despite the limitations, we believe our measures of welfare loss are more meaningful economically and nearly as easy to use as the measures currently in use to evaluate the *ex ante* performance of plan payment systems. Although developed here in the context of the ACA Marketplaces, the metrics are designed to be applicable to other private health insurance markets in the U.S. Medicare system, and in other countries relying on risk adjustment tied to other plan payment features to address issues of adverse selection in individual health insurance.

References

Akerlof G.A. (1970), "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics* 84(3): 488-500.

Beck, K., M. Trottman and P. Zweifel (2010), "Risk adjustment in Health Insurance and Its Long-Term Consequences," *Journal of Health Economics* 29(4): 489-98.

Beck K, Buchner F, van Kleef R, von Wyl V. (2014) "Risk equalization and cost saving options: How to marry solidarity with efficiency?" Unpublished.

Breyer, F., M. Heineck and N. Lorenz (2003) "Determinants of Health Care Utilization by German Sickness Fund Members – with Application to Risk Adjustment," *Health Economics* 12(5): 367-76.

Breyer F., Bundorf K., and Pauly M.V. (2012) "Health Care Spending Risk, Health Insurance, and Payment to Health Plans," in M. Pauly, T. McGuire, P. Barros (eds.) *The Handbook of Health Economics, Vol. 2,* Elsevier.

Brown J, Duggan M, Kuziemko I, and Woolston W. (2014) "How does risk selection respond to risk adjustment? Evidence from the Medicare Advantage program," *American Economic Review* 104(10): 3335-3364.

Bundorf, M. K, J. D. Levin and N. Mahoney. (2012) "Pricing and Welfare in Health Plan Choice," *American Economic Review* 102(7): 3214-3248.

Cao, Z. and McGuire, T. 2003, "Service-Level Selection by HMOs in Medicare," *Journal of Health Economics* 22(6): 915-931.

Carey, C. (2014a) "Government Payments and Insurer Benefit Design In Medicare Part D," Working Paper.

Carey, C. (2014b) "Sharing the Burden of Subsidization: Evidence from a Payment Update in Medicare Part D," Working Paper.

Chetty, R. (2009) "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods," *Annual Review of Economics*, 1:451-488.

Cohen and Einav, (2007) "Estimating Risk Preferences from Deductible Choice," *American Economic Review*, 97(3): 745-788.

Cutler D.M. and Reber S.J. (1998) "Paying for Health Insurance: The Tradeoff between Competition and Adverse Selection," *The Quarterly Journal of Economics* 113(2): 433-466.

Cutler D.M. and Zeckhauser R. (2000) "The Anatomy of Health Insurance," *Handbook of Health Economics*. Culyer A.J. and Newhouse J.P. (eds) Amsterdam: North Holland, Elsevier.

Eggleston K. and Bir A. (2009), “Measuring Selection Incentives in Managed Care: Evidence from the Massachusetts State Employees Insurance Program,” *Journal of Risk and Insurance* 76: 159-175.

Einav L. and Finkelstein A. (2011) “Selection in Insurance Markets: Theory and Empirics in Pictures,” *Journal of Economic Perspectives* 25(1): 115-138.

Einav L., Finkelstein A. and Cullen M.R. (2010) “Estimating Welfare in Insurance Markets Using Variation in Prices,” *The Quarterly Journal of Economics* 125(3): 877-921.

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.

Einav, L., Finkelstein A. and Levin, J. (2010) “Beyond Testing: Empirical Models of Insurance Markets,” *Annual Review of Economics* 2:311-36.

Ellis, R.P., Jiang S. and Kuo, T-C. (2013) “Does Service-Level Spending Show Evidence of Selection Across Health Plan Types?” *Applied Economics* 45(13): 1701-12.

Ellis R.P. and McGuire, T.G. (2007) “Predictability and Predictiveness in Health Care Spending,” *Journal of Health Economics* 26(1): 25-48.

Ericson, K.M. and Starc, A. (2012), “Age-Based Heterogeneity and Pricing Regulation on the Massachusetts Health Insurance Exchange.” NBER Working Paper 18089.

Ettner, S., Frank, R., McGuire, T. and Hermann, R. (2001), “Risk Adjustment Alternatives in Paying for Behavioral Health Care Under Medicaid,” *Health Services Research* 36(4): 793-811.

Frank R.G., Glazer J. and McGuire T.G. (2000) “Measuring Adverse Selection in Managed Health Care,” *Journal of Health Economics* 19(6): 829-854.

Geruso M. (2014) “Selection in Employer Health Plans: Homogeneous Prices and Heterogeneous Preferences”, Unpublished.

Geruso M. and Layton, T. (2015) “Upcoding: Evidence from Medicare on Squishy Risk Adjustment,” NBER Working Paper 21222.

Geruso, M. and McGuire, T.G. (2015) "Tradeoffs in the Design of Health Plan Payment Systems: Fit, Power and Balance," NBER Working Paper 20359 (revised).

Glazer J. and McGuire T.G. (2000), “Optimal Risk Adjustment of Health Insurance Premiums: an Application to Managed Care,” *The American Economic Review* 90(4): 1055-1071.

Glazer J. and McGuire T.G. (2002) “Setting Health Plan Premiums to Ensure Efficient Quality in Health Care: Minimum Variance Optimal Risk Adjustment,” *Journal of Public Economics* 84(2): 153-173.

Hackmann, M.B., Kolstad J.T., and Kowalski A.E. 2015. "Adverse Selection and an Individual Mandate: When Theory Meets Practice." *American Economic Review*, 105(3): 1030-66.

Handel B. (2013), "Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts," *American Economic Review* 103(7): 2643-2682.

Handel B., and Kolstad J.T. 2015. "Health Insurance for "Humans": Information Frictions, Plan Choice, and Consumer Welfare." *American Economic Review*, 105(8): 2449-2500.

Handel B., Hendel I. and Whinston M.D. (2015), "Equilibria in Health Exchanges: Adverse Selection vs. Reclassification Risk," *Econometrica* 83(4): 1261-1313.

HealthConnector (2013) "Commonwealth of Massachusetts Notice of Benefit and Payment Parameters 2014: Risk Adjustment Methodology and Operation." (Accessed March 23, 2015) Available at: https://betterhealthconnector.com/wp-content/uploads/reports-and-publications/Risk_Adjustment/MANoticeofBenefitPaymentParameters.pdf

Hsu J., Fung V., Huang J., et al. (2010), "Fixing Flaws in Medicare Drug Coverage that Prompt Insurers to Avoid Low-Income Patients." *Health Affairs*, vol. 29(12): 119-141.

Kautter, J., Pope, G.C., Ingber, M., Freeman, S., Patterson, L., Cohen, M. and Keenan, P. (2014), "The HHS-HCC Risk Adjustment Model for Individual and Small Group Markets under the Affordable Care Act," *Medicare & Medicaid Research Review* 4(3): E1-E11.

Kronick, R. and Welch, P.W. (2014), "Measuring Coding Intensity in the Medicare Advantage Program," *Medicare & Medicaid Research Review* 4(2): E1-E19.

Keeler, E., Carter, G. and Newhouse, J. (1998), "A Model of the Impact of Reimbursement Schemes on Health Plan Choice," *Journal of Health Economics* 17(3): 297-320.

Kuziemko I, Meckel K, Rossin-Slater M. (2014) "Do Insurers Risk-Select Against Each Other? Evidence from Medicaid and Implications for Health Reform." NBER Working Paper No. 19198.

Kowlaski, A. 2014 "The Early Impact of the Affordable Care Act, State by State," Brookings Papers on Economic Activity, Economic Studies Program, The Brookings Institution, vol. 49(2 (Fall)), pages 277-355.

Layton, T., (2014), "Imperfect Risk Adjustment, Risk Preferences, and Sorting in Competitive Health Insurance Markets," Unpublished.

Layton, T., McGuire, T.G. and Sinaiko, A.S. (forthcoming), "Risk Corridors and Reinsurance in Health Insurance Marketplaces: Insurance for Insurers" *American Journal of Health Economics*.

Lorenz, N. (2014) "Using Quantile Regression for Optimal Risk Adjustment." Universitat Trier Research Papers in Economics No. 2014-11. http://www.uni-trier.de/fileadmin/fb4/prof/VWL/EWF/Research_Papers/2014-11.pdf.

McGuire, T. G., Glazer, J., Newhouse, J.P., Normand, S-L., Shi, J., Sinaiko, A.D. Zuvekas, S. (2013) "Integrating Risk Adjustment and Enrollee Premiums in Health Plan Payment," *Journal of Health Economics* 32(6): 1263-1277.

McGuire, T.G., Newhouse, J.P., Normand, S-L., Shi, J. and Zuvekas, S. (2014) "Assessing Incentives for Service-Level Selection in Private Health Insurance Exchanges," *Journal of Health Economics* 35: 47-63.

Newhouse, J.P., W.G. Manning, E.B. Keeler and E.M. Sloss (1989), "Adjusting Capitation Rates using Objective Health Measures and Prior Utilization," *Health Care Financing Review* 15(1): 39-54.

Newhouse JP, McWilliams JM, Price M, Huang J, Fireman B, Hsu J, (2013) "Do Medicare Advantage Plans Select Enrollees in Higher Margin Clinical Categories?" *Journal of Health Economics* 32(6): 1278-1288.

Newhouse JP, Price M, McWilliams JM, Hsu J, McGuire TG. (2015) "How Much Favorable Selection Is Left in Medicare Advantage?" *American Journal of Health Economics*, 1(1): 1-26.

Pauly M. (2008) "Adverse Selection and Moral Hazard: Implications for Health Insurance Markets," *Incentives and Choice in Health Care*, Sloan F. and Kasper, H. (eds) Cambridge, MA: MIT Press, 2008

Pope, G.C., Bachofer, H., Pearlman, A., Kautter, J., Hunter, E., Miller, D. and Keenan, P. (2014) "Risk Transfer Formula for Individual and Small Group Markets Under the Affordable Care Act," *Medicare & Medicaid Research Review* 4(3): E1-E23.

Pope, G. C., Kautter, J., Ingber, J.J., Freeman, S., Sekar, R., Newhart, C. (2011), Evaluation of the CMS-HCC Risk Adjustment Model, In Final Report, RTI Project Number 0209853.006, RTI International, March.

Rothschild M and Stiglitz J. 1976. "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information." *The Quarterly Journal of Economics* 90(4): 629-649.

Rose, S., Shi, J., McGuire, T.G. and Normand, S-L. (2015) "Matching and Imputation Methods for Risk Adjustment in Health Insurance Exchanges," *Statistics in Biomedical Sciences* (Special Issue on Big Data). Advance online publication.

Shepard, M. (2015) "Hospital Network Competition and Adverse Selection: Evidence from the Massachusetts Health Insurance Exchange," unpublished.

Shmueli, A., Messika, D., Zmora, I. and Oberman, B. (2010) "Health Care Costs During the Last 12 Months of Life in Israel: Estimation and Implications for Risk Adjustment," *International Journal of Health Care Finance and Economics* 10(3): 257-73.

Van Barneveld, E., L. Lamers, R. Van Vliet and W. Van de Ven (2001), "Risk Sharing as a Supplement to Imperfect Capitation: A Tradeoff Between Selection and Efficiency," *Journal of Health Economics* 20(2): 147-168.

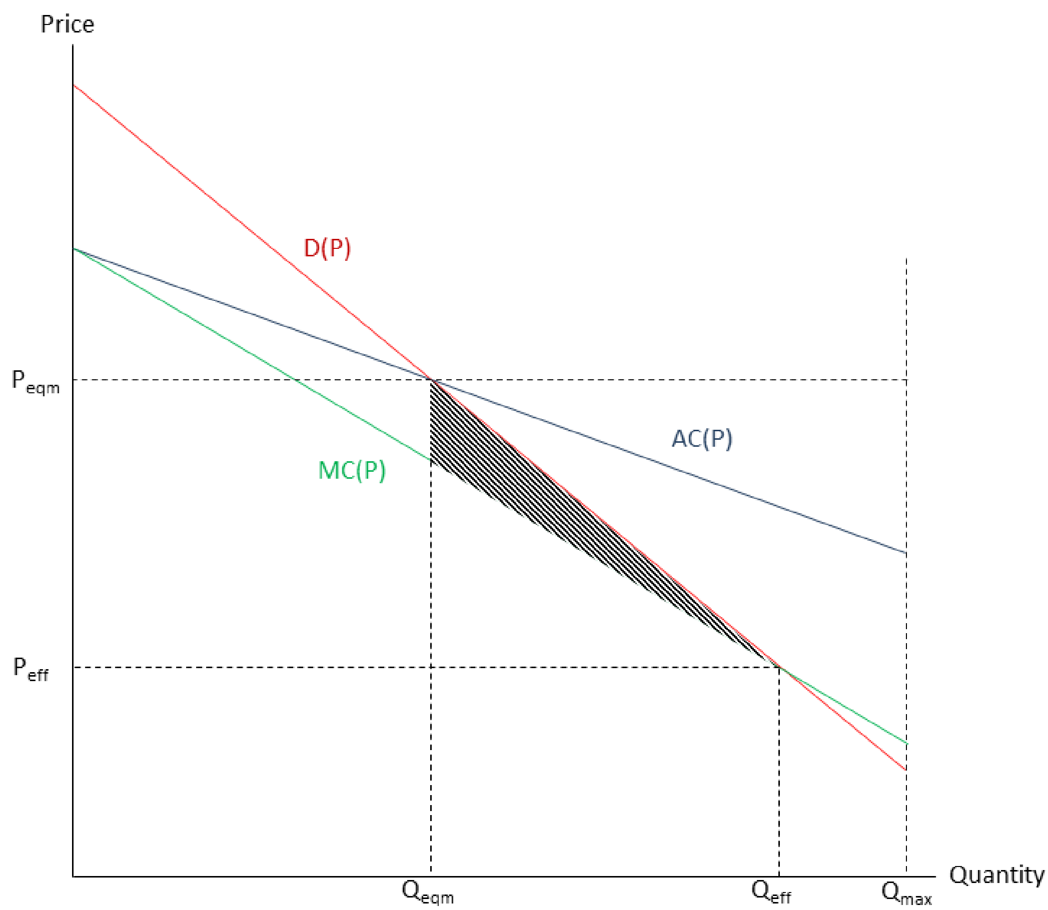
van de Ven, W.P.M.M. and Ellis, R.P. (2000) "Risk Adjustment in Competitive Health Plan Markets," *Handbook of Health Economics* Culyer, A.J. and Newhouse, J.P. (eds.) Amsterdam, Elsevier.

van Veen, S.H.C.M., van Kleef, R.C., van de Ven, W.P.M.M. and van Vliet, R.C.J.A. (2015) "Is There One Measure of Fit that Fits All? A Taxonomy and Review of Measures of Fit for Risk Equalization Models," *Medical Care Research and Review*, pp 1-24.

van Kleef, R.C., van Vliet, R.C.J.A. and van de Ven, W.P.M.M. (2013) "Risk Equalization in The Netherlands: an Empirical Evaluation," *Expert Review of Pharmacoeconomic Outcomes Research* 13(6): 829-39.

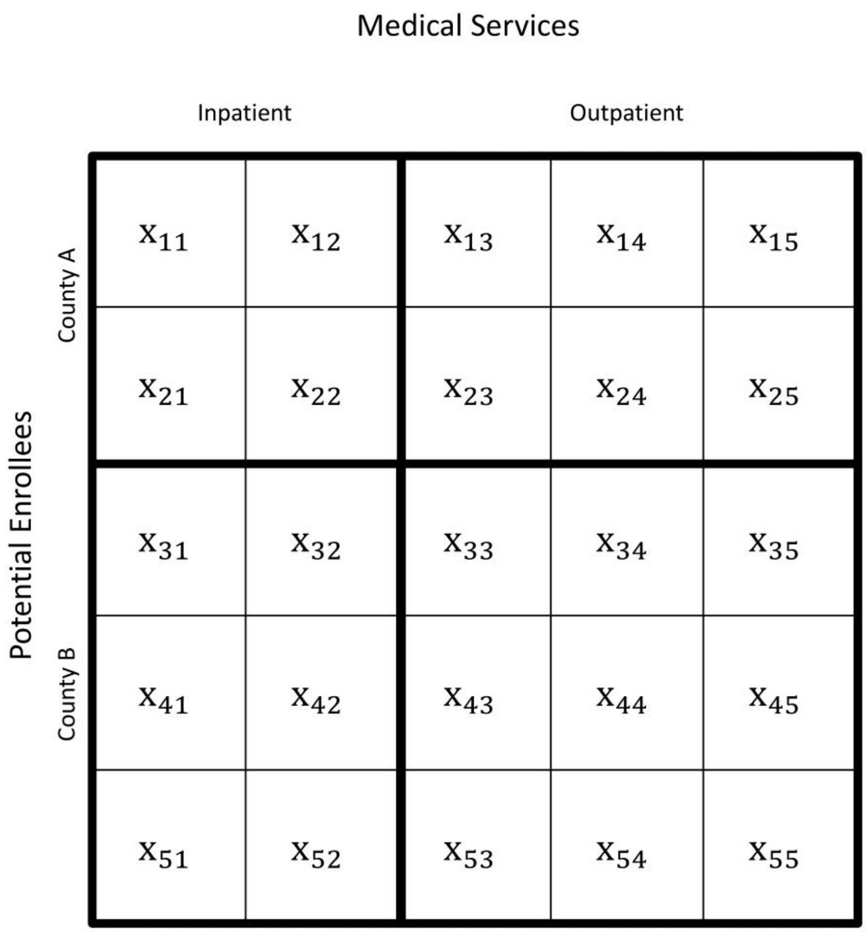
Zhu, J.M., Layton, T.J., Sinaiko, A.D. and McGuire T.G. (2013) "The Power of Reinsurance in Health Insurance Exchanges to Improve the Fit of the Payment system and Reduce Incentives for Adverse Selection," *Inquiry* 50(4): 255-75.

Figure 1. Efficiency Costs of Adverse Selection in the Einav-Finkelstein Model



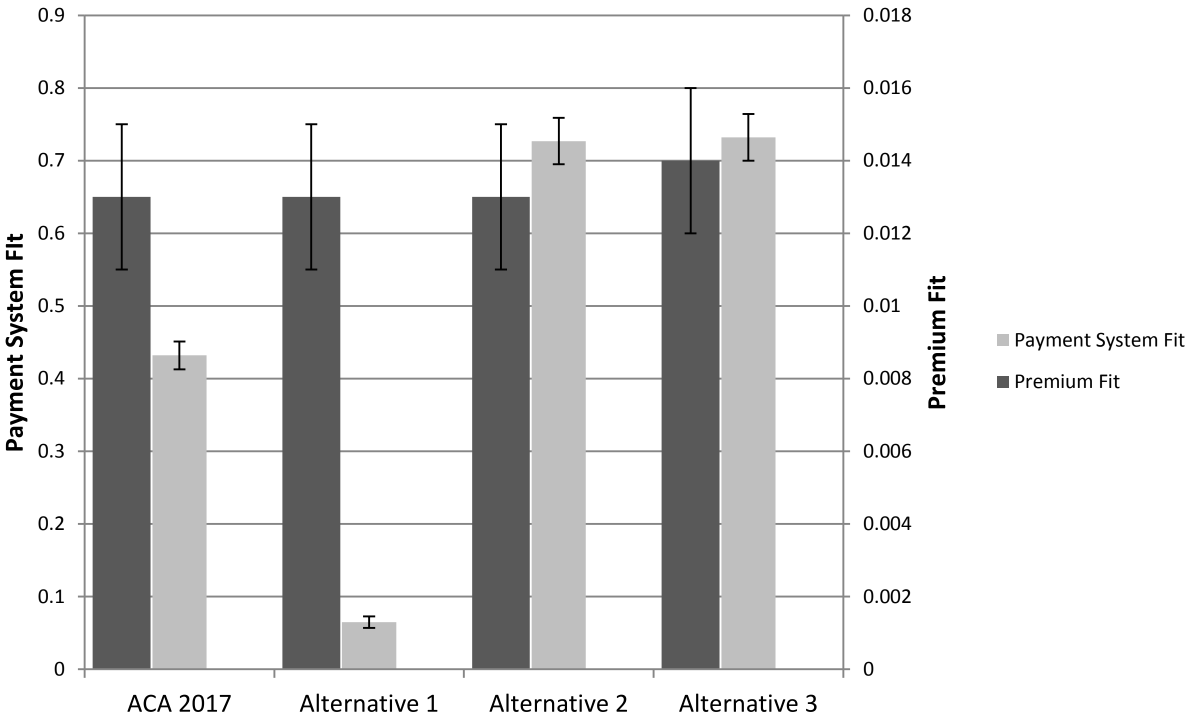
Source: Einav, Finkelstein, and Cullen (2010)

Figure 2. Possible Partition of Medical Services Spending by Person and Service



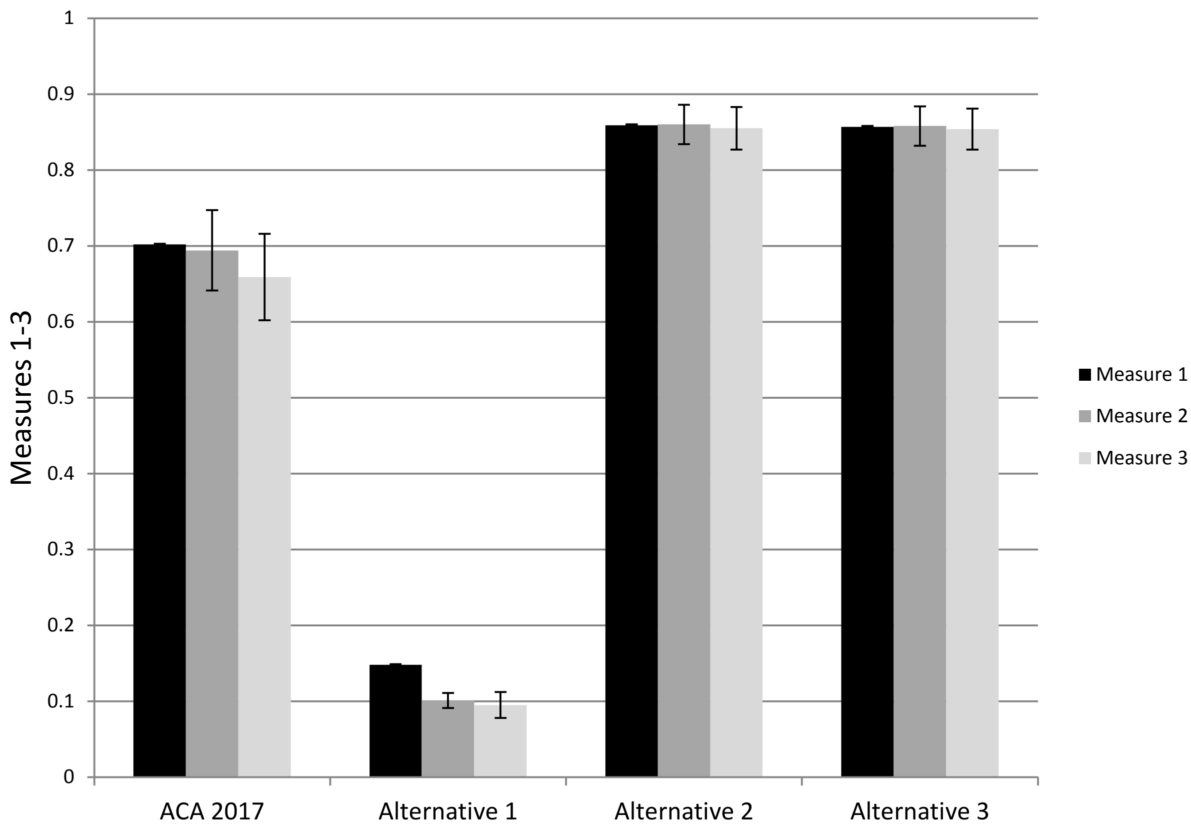
Notes: The figure illustrates the partitioning of total medical spending under an insurance contract based on services and individuals. There are 5 services and 5 individuals. The researcher must choose on which service- or person-level groupings insurers can discriminate. The figure illustrates a case where insurers can discriminate on county and inpatient/outpatient spending, resulting in 4 cells across which the insurer can distort spending allocations to attract (avoid) the healthy (sick).

Figure 3. Premium Fit and Payment System Fit for Each Simulated Payment System



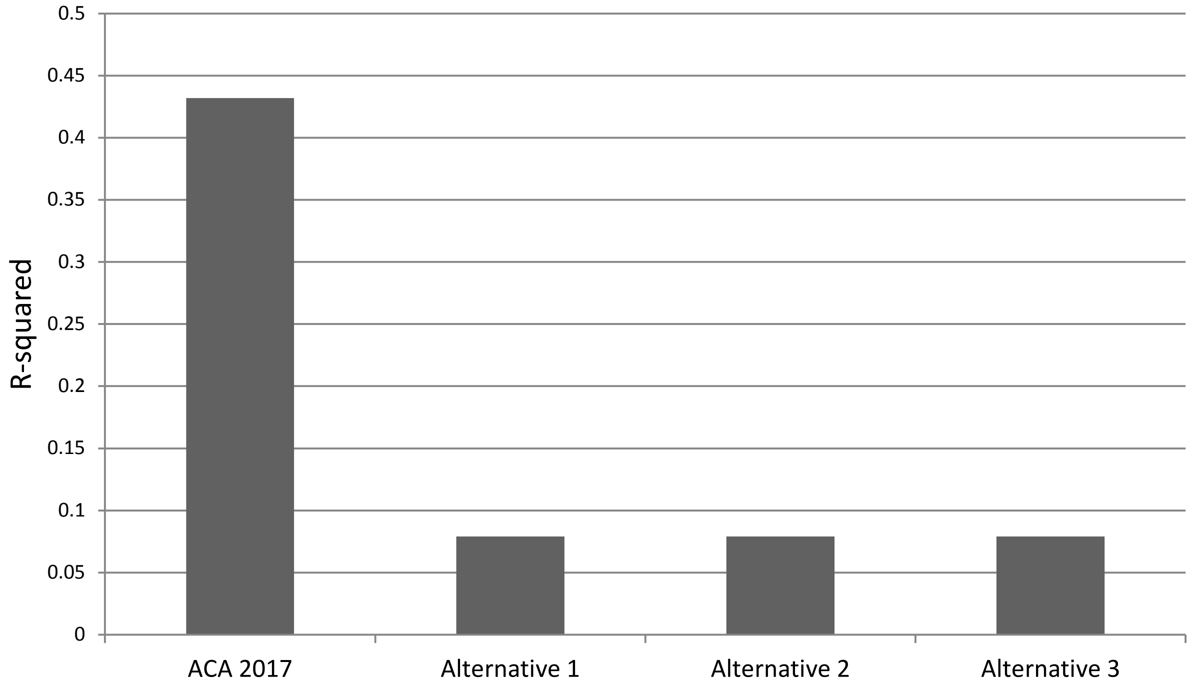
Notes: The figure displays premium fit (right y-axis) and payment system fit (left y-axis), the measures of welfare loss due to price distortions caused by premium regulations and adverse selection derived in Section 3. 1 implies complete elimination of the inefficiency and 0 implies reduction in inefficiency. Premium fit is the portion of the variance in costs explained by premium categories. Payment system fit is the portion of the variance in costs explained by variation in transfers and premiums. Error bars show bootstrapped 90% confidence intervals. The simulated payment systems are as follows: ACA 2017, concurrent risk adjustment + federal age curve; Alternative 1, Prospective risk adjustment (w/ 50% "new enrollees" assigned age/gender risk scores) + federal age curve; Alternative 2, Prospective risk adjustment (w/ 50% "new enrollees" assigned age/gender risk scores) + federal age curve + reinsurance; Alternative 3, Prospective risk adjustment (w/ 50% "new enrollees" assigned age/gender risk scores) + reinsurance + freely set age-based premiums.

Figure 4. Measures of Removal of Insurer Incentives to Distort Health Plan Benefits (ψ)



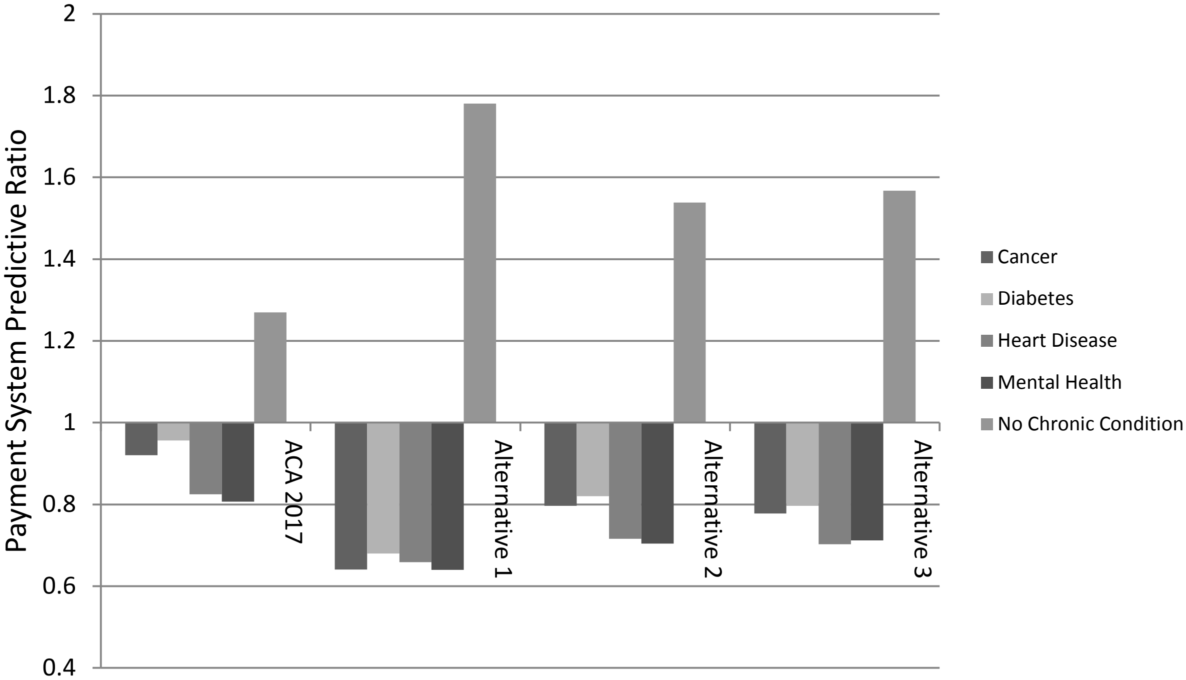
Notes: The figure displays the measure of inefficiency due to insurer incentives to distort health plan benefits due to adverse selection (ψ), derived in Section 4 of the paper. The measure describes inefficiency of the given payment system relative to a base payment system with no transfers and a single premium. 1 implies complete elimination of the inefficiency and 0 implies no better than the base payment system. ψ_1-ψ_3 are three variants on ψ. ψ_1 assumes insurers can independently discriminate against individuals with any of four chronic diseases: Cancer, heart disease, diabetes, and mental health conditions. ψ_2 assumes insurers can discriminate only at the level of broad service categories: inpatient, outpatient, and prescription drugs. ψ_3 assumes insurers can discriminate at the service level and by age (over 40/under 40). The simulated payment systems are as follows: ACA 2017, concurrent risk adjustment + federal age curve; Alternative 1, Prospective risk adjustment (w/ 50% "new enrollees" assigned age/gender risk scores) + federal age curve; Alternative 2, Prospective risk adjustment (w/ 50% "new enrollees" assigned age/gender risk scores) + federal age curve + reinsurance; Alternative 3, Prospective risk adjustment (w/ 50% "new enrollees" assigned age/gender risk scores) + reinsurance + freely set age-based premiums

Figure 5. R-squared Statistics for the Payment System Alternatives



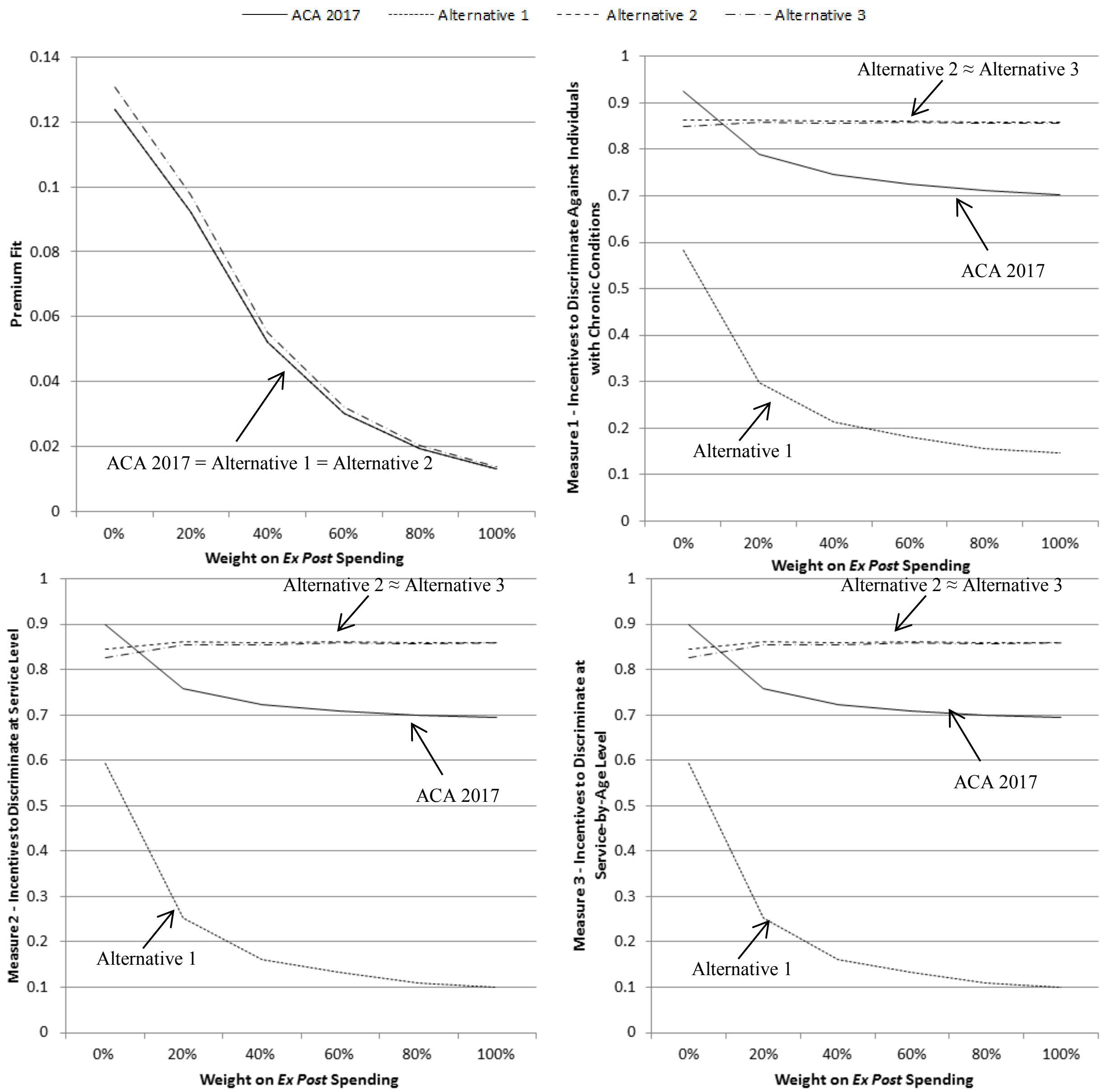
Notes: The figure presents the conventional measure of the performance or fit of a risk adjusted payment system. The R-squared measure is the R-squared statistic from a regression of costs on risk scores from the risk adjustment model used in the given payment system. The simulated payment systems are as follows: ACA 2017, concurrent risk adjustment + federal age curve; Alternative 1, Prospective risk adjustment (w/ 50% "new enrollees" assigned age/gender risk scores) + federal age curve; Alternative 2, Prospective risk adjustment (w/ 50% "new enrollees" assigned age/gender risk scores) + federal age curve + reinsurance; Alternative 3, Prospective risk adjustment (w/ 50% "new enrollees" assigned age/gender risk scores) + reinsurance + freely set age-based premiums

Figure 6. Payment System Predictive Ratios



Notes: The figure presents payment system predictive ratios for various disease groups under each of the simulated payment systems. Payment system predictive ratios are defined as average revenues divided by average costs for the specified group under the given payment system. Revenues incorporate both risk adjustment and reinsurance. Conventional predictive ratios would be identical to payment system predictive ratios for the ACA 2017 payment system. For Alternatives 1-3, conventional predictive ratios would be equal to the payment system predictive ratios under Alternative 1, as this payment system consists only of risk adjustment. The chronic disease groups are mutually exclusive. If an individual has one diagnosis mapping to a CCS code corresponding to the chronic disease, she is assigned to that chronic disease group. If she is assigned to multiple chronic disease groups, she is assigned only to the one for which she has the highest spending. The simulated payment systems are as follows: ACA 2017, concurrent risk adjustment + federal age curve; Alternative 1, Prospective risk adjustment (w/ 50% “new enrollees” assigned age/gender risk scores) + federal age curve; Alternative 2, Prospective risk adjustment (w/ 50% “new enrollees” assigned age/gender risk scores) + federal age curve + reinsurance; Alternative 3, Prospective risk adjustment (w/ 50% “new enrollees” assigned age/gender risk scores) + reinsurance + freely set age-based premiums

Figure 7. Sensitivity of Selection Measures to Assumptions about Consumer Expectations



Notes: Figures show how selection measures vary as the weight on *ex post* spending relative to predicted spending increases. Predicted spending is specified as the predicted value from a regression of 2013 spending on age-by-sex cells and deciles of 2012 inpatient, outpatient, and prescription drug spending. Each figure shows a different selection measure, labeled on the y-axis. For all measures, 1 is interpreted as total elimination of inefficiency and 0 is interpreted as no reduction in inefficiency. The simulated payment systems are as follows: ACA 2017, concurrent risk adjustment + federal age curve; Alternative 1, Prospective risk adjustment (w/ 50% "new enrollees" assigned age/gender risk scores) + federal age curve; Alternative 2, Prospective risk adjustment (w/ 50% "new enrollees" assigned age/gender risk scores) + federal age curve + reinsurance; Alternative 3, Prospective risk adjustment (w/ 50% "new enrollees" assigned age/gender risk scores) + reinsurance + freely set age-based premiums

Table 1: Special Cases of Measure of Incentives to Inefficiently Distort Health Plan Benefits

	No Service-level Selection	Yes Service-level Selection
No Individual-level Selection	$\psi = 1 - \frac{\left(\sum_i \frac{x_i}{x}\left(rev_{p,i} - x_i\right)\right)^2}{\left(\sum_i \frac{x_i}{x}(\bar{x} - x_i)\right)^2}$	$\psi = 1 - \frac{\sum_s \left(\sum_i \frac{x_{is}}{x_s}\left(rev_{p,i} - x_i\right)\right)^2}{\sum_s \left(\sum_i \frac{x_{is}}{x_s}(\bar{x} - x_i)\right)^2}$
Yes Individual-level Selection	$\psi = 1 - \frac{\sum_i \left(rev_{p,i} - x_i\right)^2}{\sum_i (\bar{x} - x_i)^2} = R^2(x, Rev_{p,i})$	$\psi = 1 - \frac{\sum_i \left(rev_{p,i} - x_i\right)^2}{\sum_i (\bar{x} - x_i)^2} = R^2(x, Rev_{p,i})$

Notes: This table shows special cases of our measure of an insurer's incentive to inefficiently distort plan benefits to attract healthy enrollees. The measure requires the researcher/policymaker to choose a particular partition of spending across people and services.

Table 2: Summary of Selection Measures

	Measure	Description
Premium Fit	$\delta = 1 - \frac{\sum_i (x_i - \bar{x}_a)^2}{\sum_i (x_i - \bar{x})^2}$	R-squared from regression of spending on a set of indicators for premium groups. Based on expected spending from social planner's point of view.
Payment System Fit	$\phi = 1 - \left[\frac{\sum_{i=1}^{N}\left(x_i - \left(r_{p,i} + \left(\bar{x}_a - \bar{r}_{p,a}\right)\right)\right)^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}\right]$	R-squared from regression producing predicted values equal to $\left(r_{p,i} + \left(\bar{x}_a - \bar{r}_{p,a}\right)\right)$. Based on expected spending from the insurer's point of view. Captures Einav-Finkelstein type inefficiency.
Incentives to Distort Benefits	$\psi = 1 - \frac{\sum_q \left(\sum_i \frac{\hat{x}_i^q}{\hat{X}^q}\left(rev_{p,i} - x_i\right)\right)^2}{\sum_q \left(\sum_i \frac{\hat{x}_i^q}{\hat{X}^q}\left(\bar{x} - x_i\right)\right)^2}$	R-squared-like measure incorporating the "predictiveness" of different services for different types of people. Revenue must be simulated for the chosen payment system and the researcher must specify a partition of spending corresponding to the margins on which the insurer can distort benefits. Based on expected spending from the insurer's point of view of the consumer's point of view. Captures "service-level" selection and other cases.

Notes: This table presents our three measures of selection-related inefficiencies developed in Sections 3 and 4. The first measure captures the inefficiency due to limited premium categories the second measure captures inefficient sorting, and the third measure captures incentives to inefficiently distort plan benefits to attract healthy enrollees.

Table 3: Data Used for Estimation and Simulation: Means and Standard Deviations

	Marketplace Sample (N=2,006,126)		Random Sample of Full MarketScan (N=498,398)	
	Mean	Std. Dev.	Mean	Std. Dev.
Age	42.40	12.46	44.70	11.58
Female	0.49		0.52	
Census Region:				
Northeast	0.14		0.17	
Central	0.23		0.26	
South	0.43		0.39	
West	0.20		0.17	
Total Spending	$5,059	$20,134	$5,838	$18,779
Inpatient Spending	$1,357	$13,657	$1,221	$10,910
Outpatient Spending	$2,779	$10,204	$3,344	$11,150
Drug Spending	$922	$4,189	$1,273	$5,251
One or More Chronic Conditions	0.33		0.40	
Cancer	0.07		0.08	
Heart Disease	0.07		0.09	
Mental Health	0.11		0.13	
Diabetes	0.08		0.11	
Recalibrated Concurrent RRS	1.0	2.63	1.0	2.02
Recalibrated Prospective RRS	1.0	1.45	1.0	1.25
Age-sex only RRS	1.0	0.49	1.0	0.36

Notes: Our Full MarketScan sample consists of adults (21-64 in 2012) who are continuously enrolled for 2012-13 and have drug and mental health coverage. Individuals must be enrolled in a plan type that does not include capitated payments to providers. The Marketplace Sample is a subset of the Full MarketScan sample, selected with propensity score methods described in the text. The recalibrated concurrent and prospective models use our Marketplace sample for calibration, and hence have means of one once normalized. Also shown are the sample standard deviations.

Table 4: Simulated Payment Systems

	Current Law	Alternative Payment Systems		
	ACA 2017	1	2	3
Premiums	Specified age curve	Specified age curve	Specified age curve	Age curve set in equilibrium
Risk Adjustment	HHS-HCC Concurrent	HHS-HCC Prospective, Age-Gender for new enrollees	HHS-HCC Prospective, Age-Gender for new enrollees	HHS-HCC Prospective, Age-Gender for new enrollees
Reinsurance	No	No	Yes	Yes
ACA Risk Equalization Transfer Formula	Yes	Yes	Yes	No: age curve not used

Table 5: Estimated Measures for Simulated Payment Systems

Measure	Payment System: ACA 2017	Alternative 1	Alternative 2	Alternative 3
Premium Fit: δ	0.013	0.013	0.013	0.014
	(.012, .014)	(.012, .014)	(.012, .014)	(.015, .013)
Payment System Fit: ϕ	0.423	0.065	0.727	0.732
	(.414, .433)	(.061, .069)	(.711, .743)	(.716, .748)
Incentives to Distort Benefits:				
ψ_1 – Four Chronic Diseases	0.702	0.148	0.859	0.857
	(.701, .702)	(.147, .148)	(.859, .860)	(.857, .858)
ψ_2 – Service-level	0.694	0.101	0.860	0.858
	(.668, .721)	(.091, .101)	(.846, .872)	(.845, .871)
ψ_3 – Service-level and Young/Old	0.659	0.095	0.855	0.854
	(.655, .712)	(.087, .104)	(.840, .868)	(.840, .867)

Notes: Shown are point estimates (and confidence intervals) for measures of the inefficiency of the given payment system. For each measure, 1 implies complete elimination of the inefficiency and 0 implies no reduction in inefficiency. Premium fit is the portion of the variance in costs explained by premium categories. Payment system fit is the portion of the variance in costs explained by variation in premiums and transfers. ψ_1-ψ_3 are three variants of our measure of insurer incentives to distort plan benefits. ψ_1 assumes insurers can independently discriminate against individuals with any of four chronic diseases: Cancer, heart disease, diabetes, and mental health conditions. ψ_2 assumes insurers can discriminate only at the level of broad service categories: inpatient, outpatient, and prescription drugs. ψ_3 assumes insurers can discriminate at the level of broad service levels and by age (over 40/under 40). The four simulated payment systems are shown: ACA 2017 is concurrent risk adjustment + federal age curve; Alternative 1 is Prospective risk adjustment (with 50% "new enrollees" assigned age/gender risk scores) + federal age curve; Alternative 2 is Prospective risk adjustment (with 50% "new enrollees" assigned age/gender risk scores) + federal age curve + reinsurance; Alternative 3 is Prospective risk adjustment (with 50% "new enrollees" assigned age/gender risk scores) + reinsurance + freely set age-based premiums.

Table 6: Conventional Measures for Simulated Payment Systems

Measure	Payment System			
	ACA 2017	Alternative 1	Alternative 2	Alternative 3
R-squared	0.432	0.079	0.079	0.079
Payment System Predictive Ratios:				
Cancer	0.920	0.641	0.797	0.778
Diabetes	0.956	0.680	0.820	0.797
Heart Disease	0.825	0.659	0.716	0.703
Mental Health	0.807	0.640	0.704	0.712
No Chronic Condition	1.270	1.780	1.538	1.567

Notes: The R-squared measure is the R-squared statistic from a regression of costs on risk scores from the risk adjustment model used in the given payment system. Predictive ratios are defined as average revenues divided by average costs for the specified group under the given payment system. The chronic disease groups are mutually exclusive. If an individual has one diagnosis mapping to a CCS code corresponding to the chronic disease, she is assigned to that chronic disease group. If she is assigned to multiple chronic disease groups, she is assigned only to the one for which she has the highest spending. The four simulated payment systems are shown: ACA 2017 is concurrent risk adjustment + federal age curve; Alternative 1 is Prospective risk adjustment (with 50% "new enrollees" assigned age/gender risk scores) + federal age curve; Alternative 2 is Prospective risk adjustment (with 50% "new enrollees" assigned age/gender risk scores) + federal age curve + reinsurance; Alternative 3 is Prospective risk adjustment (with 50% "new enrollees" assigned age/gender risk scores) + reinsurance + freely set age-based premiums.

Appendix A: Derivations from Section 3

A.i. Derivation of Equation (2)

This appendix includes the derivations of the expressions/measures of sorting inefficiencies due to premium regulations and adverse selection presented in Section 3. We start by deriving Equation (1) from the text. Let x_i be the (expected) cost to an insurer of enrolling person i in a basic ("silver") plan and let $(1 + \gamma)x_i$ be the cost to an insurer of enrolling person i in a generous ("gold") plan, with $\gamma > 0$. Person i's expected incremental cost is defined as the difference between her expected plan cost in the gold plan and her expected plan cost in the silver plan: $(1+\gamma)x_i - x_i = \gamma x_i$. Consumers fall into T types, where all type-t consumers have the same incremental expected cost: $\gamma x_i = \gamma x_t$ for all $i \in t$. Finally, let prem_i^j be the premium the insurer charges person i to enroll in plan j, j = g,s, and let the incremental premium charged to person i, be $\mathrm{prem}_i = \mathrm{prem}_i^g - \mathrm{prem}_i^s$.

Consumers have heterogeneous preferences for plans, but we require all consumers to choose either the gold or the silver plan. Let $n_t(\mathrm{prem}_t)$ be the number of type-t consumers who purchase gold given the incremental price, prem_t, and let $P_t(n)$ be the "inverse demand," or the price at which n type-t consumers enroll in gold. In this setting, "incremental" social welfare for type-t consumers can be described as a function of prem_t:

$$W_t(\mathrm{prem}_t) = \left(\int_0^{n_t(\mathrm{prem}_t)} P_t(n)dn \right) - n_t(\mathrm{prem}_t)\gamma x_t$$

This function is maximized at $\mathrm{prem}_t^* = \gamma x_t$. The welfare loss due to a price different from $\mathrm{prem}_t^* = \gamma x_t$ can be expressed as $\Delta W_t(\mathrm{prem}_t) = W_t(\gamma x_t) - W_t(\mathrm{prem}_t)$:

$$\Delta W_t(\mathrm{prem}_t) = \left[\left(\int_0^{n_t(\gamma x_t)} P_t(n)dn \right) - n_t(\gamma x_t)\gamma x_t \right] - \left[\left(\int_0^{n_t(\mathrm{prem}_t)} P_t(n)dn \right) - n_t(\mathrm{prem}_t)\gamma x_t \right]$$

Now, take a 2nd order Taylor series expansion of $\Delta W_t(\mathrm{prem}_t)$ around $\mathrm{prem}_t = \gamma x_t$,

$$\Delta W_t(\mathrm{prem}_t) \approx \Delta W_t(\gamma x_t) + \frac{\partial \Delta W_t(\gamma x_t)}{\partial \mathrm{prem}_t}(\mathrm{prem}_t - \gamma x_t) + \frac{1}{2}\frac{\partial^2 \Delta W_t\ \ (\gamma x_t)}{\partial \mathrm{prem}_t^2}(\mathrm{prem}_t - \gamma x_t)^2$$

$$\Delta W_t(\mathrm{prem}_t) \approx -\{[p_t n_t'(\mathrm{prem}_t) - n_t'(\mathrm{prem}_t)\gamma x_t]\}_{\mathrm{prem}_t = \gamma x_t}(\mathrm{prem}_t - \gamma x_t)$$
$$- \frac{1}{2}\{\mathrm{prem}_t n_t''(\mathrm{prem}_t) + n_t'(\mathrm{prem}_t) - n_t''(\mathrm{prem}_t)\gamma x_t\}_{\mathrm{prem}_t = \gamma x_t}(\mathrm{prem}_t - \gamma x_t)^2$$

$$\Delta W_t(\mathrm{prem}_t) \approx -\frac{1}{2} n_t'(\gamma x_t)(\mathrm{prem}_t - \gamma x_t)^2 \tag{A1}$$

Now, sum Equation (A1) over all types,

$$\sum_{t=1}^{T} \Delta W_t(\mathrm{prem}_t) \approx -\frac{1}{2}\sum_{t=1}^{T} n_t'(\gamma x_t)(\mathrm{prem}_t - \gamma x_t)^2$$

and assume that $\varepsilon = \frac{n'_t}{N_t}$ for all t

$$\sum_{t=1}^{T} \Delta W_t(\mathrm{prem}_t) \approx -\frac{\varepsilon}{2}\sum_{t=1}^{T} N_t(\mathrm{prem}_t - \gamma x_t)^2.$$

Now, since all individuals of the same type face the same premium and have the same expected costs, this implies that:

$$\sum_{t=1}^{T} \Delta W_t(prem_t) \approx -\frac{\varepsilon}{2}\sum_{i=1}^{N}(prem_i - \gamma x_i)^2 \tag{A2}$$

A.ii. Expression for ω

In Section 3.4, we claim that the following equality holds given a set of regularity assumptions:

$$\widetilde{\phi} = \frac{\sum_{a=1}^{A} N_a \sum_{i\in a}\left((\gamma+1)\bar{x}_{p,a}^{g} - \bar{x}_{p,a}^{s} - \gamma\bar{x}_a\right)^2}{\gamma^2\sum_{i=1}^{N}(x_i-\bar{x})^2} = \omega\left[\frac{\sum_{i=1}^{N}\left(x_i - \left(r_{p,i} + \left(\bar{x}_a - \bar{r}_{p,a}\right)\right)\right)^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2}\right]$$

The regularity assumptions are as follows:

ASSUMPTION 1 (UNIFORM SELECTION): Under any payment system and within each premium category the relationship between the average cost in the gold plan and the average cost overall is similar. Formally, re-writing $\left(\bar{x}_{p,a}^{g} - \bar{x}_a\right)$ as $\eta_{p,a}\left(\bar{x}_{p,a}^{g} - \bar{x}_{p,a}^{s}\right)$, we assume that $\eta_{p,a} = \eta$ for all relevant payment systems and premium categorizations.

ASSUMPTION 2 (PROPORTIONAL SORTING): In expressions for the incremental average cost, we assume that the sorting of consumers across plans is similar across payment systems and premium groups. Write gold and silver plan enrollment among individuals in premium category a under payment system p as

$$N_{pa}^{g} = \tau_{pa}N_a$$
$$N_{pa}^{s} = (1-\tau_{pa})N_a$$

Formally, we assume that $\tau_{pa} = \tau$ for all premium groups and payment systems.

ASSUMPTION 3 (PROPORTIONAL VARIANCE): We assume that the ratio of the within-plan variance within a premium category to the total variance within a premium category is constant across premium categories. For payment system p, premium group a, and plan-type j, let ξ_{pq}^{j} be the ratio:

$$\frac{\sum_{N_a^j}\left(x_{pi} - \bar{x}_{pa}\right)^2}{\sum_{N_a}\left(x_{pi} - \bar{x}_{pa}\right)^2} = \xi_{pa}^{j}$$

Formally, we assume that for each plan type ξ_{pa}^{j} is constant across payment systems and premium groups, $\xi_{pa}^{j} = \xi^{j}$.

We will now prove the claim given the assumptions. We start by re-writing $\frac{\sum_{a=1}^{A} N_a\left((\gamma+1)\bar{x}_{p,a}^{g} - \bar{x}_{p,a}^{s} - \gamma\bar{x}_a\right)^2}{\gamma^2\sum_{i=1}^{N}(x_i-\bar{x})^2}$ as

$$\frac{\sum_{a=1}^{A} N_a\left((\gamma+1)\bar{x}_{p,a}^{g} - \bar{x}_{p,a}^{s} - \gamma\bar{x}_a\right)^2}{\gamma^2\sum_{i=1}^{N}(x_i-\bar{x})^2} = \frac{\sum_{a=1}^{A} N_a\left(\gamma\left(\bar{x}_{p,a}^{g} - \bar{x}_a\right) + \left(\bar{x}_{p,a}^{g} - \bar{x}_{p,a}^{s}\right)\right)^2}{\gamma^2\sum_{i=1}^{N}(x_i-\bar{x})^2} \tag{A3}$$

Now, if we re-write $\left(\bar{x}_{p,a}^{g} - \bar{x}_a\right)$ as $\eta_{p,a}\left(\bar{x}_{p,a}^{g} - \bar{x}_{p,a}^{s}\right)$ we can write Equation A3 as

$$\frac{\sum_{a=1}^{A} N_a\left((\gamma+1)\bar{x}_{p,a}^{g} - \bar{x}_{p,a}^{s} - \gamma\bar{x}_a\right)^2}{\gamma^2\sum_{i=1}^{N}(x_i-\bar{x})^2} = \frac{\sum_{a=1}^{A} N_a\left[\gamma^2\eta_{p,a}^2 + 1 + 2\gamma\eta_{p,a}\right]\left(\bar{x}_{p,a}^{g} - \bar{x}_{p,a}^{s}\right)^2}{\gamma^2\sum_{i=1}^{N}(x_i-\bar{x})^2}. \tag{A4}$$

Now, if we invoke Assumption 1 from above that $\eta_{p,a} = \eta$ for all premium groups a and all (relevant) payment systems p, then we can re-write Equation A4 as

$$\frac{\sum_{a=1}^{A} N_a \left((\gamma+1)\bar{x}_{p,a}^{g} - \bar{x}_{p,a}^{s} - \gamma \bar{x}_a\right)^2}{\gamma^2 \sum_{i=1}^{N} (x_i - \bar{x})^2} = \left[\eta^2 + \frac{1}{\gamma^2} + \frac{2\eta}{\gamma}\right] \frac{\sum_{a=1}^{A} N_a \left(\bar{x}_{p,a}^{g} - \bar{x}_{p,a}^{s}\right)^2}{\sum_{i=1}^{N} (x_i - \bar{x})^2}. \tag{A5}$$

We now focus our attention on the numerator of (A5), $\sum_{a=1}^{A} N_a \left(\bar{x}_{p,a}^{g} - \bar{x}_{p,a}^{s}\right)^2$. The intuition behind this expression is that for each individual the welfare loss is proportional to the square of the equilibrium premium faced by that individual, which, under our assumptions, is equal to the incremental average cost among all people in that individual's premium group/risk pool. Summing over individuals gives the expression.

Now, we will derive an expression for $\sum_{a=1}^{A} N_a \left(\bar{x}_{p,a}^{g} - \bar{x}_{p,a}^{s}\right)^2$ that consists only of values available in the claims data and variables that are constant across payment systems. We start by deriving an expression for the incremental average cost faced by individual i in premium group a. To do this, note that the total sum of squares for x_p among all members of premium group a can be written as within-plan components and across-plan components as follows:

$$SST_a(x_p) = \sum_{N_a} (x_{pi} - \bar{x}_a)^2 = \sum_{N_a^g} (x_{pi} - \bar{x}_{pa}^{g})^2 + \sum_{N_a^s} (x_{pi} - \bar{x}_{pa}^{s})^2 + N_a^g (\bar{x}_{pa}^{g} - \bar{x}_{pa})^2 + N_a^s (\bar{x}_{pa}^{s} - \bar{x}_{pa})^2 \tag{A6}$$

It is straightforward to show that

$$N_a^g (\bar{x}_{pa}^{g} - \bar{x}_{pa})^2 = N_a^g \left(\bar{x}_{pa}^{g} - \frac{N_a^s \bar{x}_{pa}^{s} + N_a^g \bar{x}_{pa}^{g}}{N_a}\right)^2$$

$$N_a^g (\bar{x}_{pa}^{g} - \bar{x}_{pa})^2 = N_a^g \left(\frac{N_a^s}{N_a}\right)^2 (\bar{x}_{p,a}^{g} - \bar{x}_{p,a}^{s})^2$$

Likewise,

$$N_a^s (\bar{x}_{pa}^{s} - \bar{x}_{pa})^2 = N_a^s \left(\frac{N_a^g}{N_a}\right)^2 (\bar{x}_{p,a}^{g} - \bar{x}_{p,a}^{s})^2$$

If we plug these expressions back into (A6), we get

$$SST_a(x_p) = \sum_{N_a^g} (x_{pi} - \bar{x}_{pa}^{g})^2 + \sum_{N_a^s} (x_{pi} - \bar{x}_{pa}^{s})^2 + N_a^g \left(\frac{N_a^s}{N_a}\right)^2 (\bar{x}_{p,a}^{g} - \bar{x}_{p,a}^{s})^2 + N_a^s \left(\frac{N_a^g}{N_a}\right)^2 (\bar{x}_{p,a}^{g} - \bar{x}_{p,a}^{s})^2 \tag{A7}$$

It is straightforward to show that

$$N_a^g \left(\frac{N_a^s}{N_a}\right)^2 (\bar{x}_{p,a}^{g} - \bar{x}_{p,a}^{s})^2 + N_a^s \left(\frac{N_a^g}{N_a}\right)^2 (\bar{x}_{p,a}^{g} - \bar{x}_{p,a}^{s})^2 = (\bar{x}_{p,a}^{g} - \bar{x}_{p,a}^{s})^2 \left[N_a^g \left(\frac{N_a^s}{N_a}\right)^2 + N_a^s \left(\frac{N_a^g}{N_a}\right)^2\right]$$

$$N_a^g \left(\frac{N_a^s}{N_a}\right)^2 (\bar{x}_{p,a}^{g} - \bar{x}_{p,a}^{s})^2 + N_a^s \left(\frac{N_a^g}{N_a}\right)^2 (\bar{x}_{p,a}^{g} - \bar{x}_{p,a}^{s})^2 = (\bar{x}_{p,a}^{g} - \bar{x}_{p,a}^{s})^2 \left[\frac{N_a^g N_a^s}{N_a}\right]$$

This allows us to simplify (A7) as follows:

$$SST_a(x_p) = \sum_{N_a^g} (x_{pi} - \bar{x}_{pa}^{g})^2 + \sum_{N_a^s} (x_{pi} - \bar{x}_{pa}^{s})^2 + (\bar{x}_{p,a}^{g} - \bar{x}_{p,a}^{s})^2 \left[\frac{N_a^g N_a^s}{N_a}\right]$$

Solving for $\left(\bar{x}^{g}_{p,a} - \bar{x}^{s}_{p,a}\right)^2$

$$\left(\bar{x}^{g}_{p,a} - \bar{x}^{s}_{p,a}\right)^2 = \left(\frac{N_a}{N_a^g N_a^s}\right)\left[\sum_{N_a}\left(x_{pi} - \bar{x}_{pa}\right)^2 - \left(\sum_{N_a^g}\left(x_{pi} - \bar{x}^{g}_{pa}\right)^2 + \sum_{N_a^s}\left(x_{pi} - \bar{x}^{s}_{pa}\right)^2\right)\right] \tag{A8}$$

Now, for premium group a and payment system p let us define δ^{g}_{pa} and δ^{s}_{pa} as the ratios of the variance of net costs among individuals choosing the gold and silver plans to the variance of net costs among all individuals in the premium group, respectively:

$$\sum_{N_a^g}\left(x_{pi} - \bar{x}^{g}_{pa}\right)^2 = \delta^{g}_{pa}\sum_{N_a}\left(x_{pi} - \bar{x}_{pa}\right)^2$$

$$\sum_{N_a^s}\left(x_{pi} - \bar{x}^{s}_{pa}\right)^2 = \delta^{s}_{pa}\sum_{N_a}\left(x_{pi} - \bar{x}_{pa}\right)^2$$

This allows us to re-write (A8) as

$$\left(\bar{x}^{g}_{p,a} - \bar{x}^{s}_{p,a}\right)^2 = \left(1 - \left(\delta^{g}_{pa} + \delta^{s}_{pa}\right)\right)\left(\frac{N_a}{N_a^g N_a^s}\right)\sum_{N_a}\left(x_{pi} - \bar{x}_{pa}\right)^2$$

Recall that this expression represents the welfare loss for person i in premium group a. Because the incremental average cost (and the welfare loss) is the same for all individuals in premium group a, we can sum this expression over all individuals in the group to get the group a welfare loss:

$$N_a\left(\bar{x}^{g}_{p,a} - \bar{x}^{s}_{p,a}\right)^2 = \left(1 - \left(\delta^{g}_{pa} + \delta^{s}_{pa}\right)\right)\left(\frac{N_a^2}{N_a^g N_a^s}\right)\sum_{N_a}\left(x_{pi} - \bar{x}_{pa}\right)^2$$

To get the total welfare loss, we need to sum over all premium groups:

$$\sum_{a=1}^{A} N_a\left(\bar{x}^{g}_{p,a} - \bar{x}^{s}_{p,a}\right)^2 = \sum_{a=1}^{A}\left(1 - \left(\delta^{g}_{pa} + \delta^{s}_{pa}\right)\right)\left(\frac{N_a^2}{N_a^g N_a^s}\right)\sum_{N_a}\left(x_{pi} - \bar{x}_{pa}\right)^2 \tag{A9}$$

Now, we invoke Assumptions 2 and 3 from above, allowing us to re-write (A9) as

$$\sum_{a=1}^{A} N_a\left(\bar{x}^{g}_{p,a} - \bar{x}^{s}_{p,a}\right)^2 = \left(1 - (\delta^{g} + \delta^{s})\right)\left[\sum_{a=1}^{A}\left(\frac{N_a^2}{\tau(1-\tau)N_a^2}\right)\sum_{N_a}\left(x_{pi} - \bar{x}_{pa}\right)^2\right]$$

The N_a^2 in the numerator and denominator cancel, allowing us to further simplify

$$\sum_{a=1}^{A} N_a\left(\bar{x}^{g}_{p,a} - \bar{x}^{s}_{p,a}\right)^2 = \left(1 - (\delta^{g} + \delta^{s})\right)\left(\frac{1}{\tau(1-\tau)}\right)\left[\sum_{a=1}^{A}\sum_{N_a}\left(x_{pi} - \bar{x}_{pa}\right)^2\right]$$

Now note that the part in brackets is now a sum over individuals. We can also plug in $x_{pi} = x_i - \left(r_{pi} - \bar{x}\right)$ and $\bar{x}_{pa} = \bar{x}_a - (\bar{r}_{p,a} - \bar{x})$ to get

$$\sum_{a=1}^{A} N_a\left(\bar{x}_{p,a}^{g} - \bar{x}_{p,a}^{s}\right)^2 = \left(1-(\delta^g+\delta^s)\right)\left(\frac{1}{\tau(1-\tau)}\right)\left[\sum_{i=1}^{N}\left(x_i-\left(r_{p,i}+\left(\bar{x}_a-\bar{r}_{p,a}\right)\right)\right)^2\right] \tag{A10}$$

Finally, we can plug (A10) back into (A5) to get that

$$\frac{\sum_{a=1}^{A} N_a\left((\gamma+1)\bar{x}_{p,a}^{g} - \bar{x}_{p,a}^{s} - \gamma\bar{x}_a\right)^2}{\gamma^2\sum_{i=1}^{N}(x_i-\bar{x})^2} = \left[\eta^2+\frac{1}{\gamma^2}+\frac{2\eta}{\gamma}\right]\frac{\sum_{a=1}^{A} N_a\left(\bar{x}_{p,a}^{g} - \bar{x}_{p,a}^{s}\right)^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2}$$

$$\frac{\sum_{a=1}^{A} N_a\sum_{i\in a}\left((\gamma+1)\bar{x}_{p,a}^{g} - \bar{x}_{p,a}^{s} - \gamma\bar{x}_a\right)^2}{\gamma^2\sum_{i=1}^{N}(x_i-\bar{x})^2}$$

$$= \left[\eta^2+\frac{1}{\gamma^2}+\frac{2\eta}{\gamma}\right]\left(1-(\delta^g+\delta^s)\right)\left(\frac{1}{\tau(1-\tau)}\right)\frac{\sum_{i=1}^{N}\left(x_i-\left(r_{p,i}+\left(\bar{x}_a-\bar{r}_{p,a}\right)\right)\right)^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2}$$

And, if we define $\omega = \left[\eta^2+\frac{1}{\gamma^2}+\frac{2\eta}{\gamma}\right]\left(1-(\delta^g+\delta^s)\right)\left(\frac{1}{\tau(1-\tau)}\right)$, which does not vary across payment systems, we get

$$\tilde{\phi} = \frac{\sum_{a=1}^{A} N_a\left((\gamma+1)\bar{x}_{p,a}^{g} - \bar{x}_{p,a}^{s} - \gamma\bar{x}_a\right)^2}{\gamma^2\sum_{i=1}^{N}(x_i-\bar{x})^2} = \omega\frac{\sum_{i=1}^{N}\left(x_i-\left(r_{p,i}+\left(\bar{x}_a-\bar{r}_{p,a}\right)\right)\right)^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2}$$

Appendix B: Adverse Selection Measures with an Age Curve

Under some payment systems, including the one in the Marketplaces, relative prices for different age groups are fixed by an age curve. Effectively, an insurer sets a single base price for plan j, $\widehat{\text{prem}}_p^j$, and then each individual is charged a price equal to the base price times an age weight, α_i: $\text{prem}_{p,i}^j = \widehat{\text{prem}}_p^j \alpha_i$.

Then the incremental base premium is equal to

$$\widehat{\text{prem}}_p = \widehat{\text{prem}}_p^g - \widehat{\text{prem}}_p^s$$

And the individual's incremental premium is equal to

$$\text{prem}_{pi} = \widehat{\text{prem}}_p \alpha_i$$

Then

$$1 - \frac{\sum_{i=1}^{N}\left(\text{prem}_{pi} - \gamma x_i\right)^2}{\sum_{i=1}^{N}(\gamma\bar{x} - \gamma x_i)^2} = 1 - \frac{\sum_{i=1}^{N}\left(\widehat{\text{prem}}_p \alpha_i - \gamma x_i\right)^2}{\sum_{i=1}^{N}(\gamma\bar{x} - \gamma x_i)^2}$$

Then, we decompose to get

$$1 - \frac{\sum_{i=1}^{N}\left(\widehat{\text{prem}}_p \alpha_i - \gamma x_i\right)^2}{\sum_{i=1}^{N}(\gamma\bar{x} - \gamma x_i)^2} = 1 - \frac{\sum_{i=1}^{N}\left(\left(\widehat{\text{prem}}_p \alpha_i - \frac{\alpha_i}{\bar{\alpha}}\gamma\bar{x}\right) + \left(\frac{\alpha_i}{\bar{\alpha}}\gamma\bar{x} - \gamma x_i\right)\right)^2}{\sum_{i=1}^{N}(\gamma\bar{x} - \gamma x_i)^2}$$

$$1 - \frac{\sum_{i=1}^{N}\left(\widehat{\text{prem}}_p \alpha_i - \gamma x_i\right)^2}{\sum_{i=1}^{N}(\gamma\bar{x} - \gamma x_i)^2} = 1 - \frac{\sum_{i=1}^{N}\left[\left(\widehat{\text{prem}}_p \alpha_i - \frac{\alpha_i}{\bar{\alpha}}\gamma\bar{x}\right)^2 + \left(\frac{\alpha_i}{\bar{\alpha}}\gamma\bar{x} - \gamma x_i\right)^2 + 2 * \left(\widehat{\text{prem}}_p \alpha_i - \frac{\alpha_i}{\bar{\alpha}}\gamma\bar{x}\right)\left(\frac{\alpha_i}{\bar{\alpha}}\gamma\bar{x} - \gamma x_i\right)\right]}{\sum_{i=1}^{N}(\gamma\bar{x} - \gamma x_i)^2}$$

Now, distribute the sum

$$1 - \frac{\sum_{i=1}^{N}\left(\widehat{\text{prem}}_p \alpha_i - \gamma x_i\right)^2}{\sum_{i=1}^{N}(\gamma\bar{x} - \gamma x_i)^2} = 1 - \frac{\sum_{i=1}^{N}\left[\left(\widehat{\text{prem}}_p \alpha_i - \frac{\alpha_i}{\bar{\alpha}}\gamma\bar{x}\right)^2 + \left(\frac{\alpha_i}{\bar{\alpha}}\gamma\bar{x} - \gamma x_i\right)^2\right] + \sum_{\alpha}\sum_{i=1}^{N_\alpha} 2 * \left(\widehat{\text{prem}}_p \alpha_i - \frac{\alpha_i}{\bar{\alpha}}\gamma\bar{x}\right)\left(\frac{\alpha_i}{\bar{\alpha}}\gamma\bar{x} - \gamma x_i\right)}{\sum_{i=1}^{N}(\gamma\bar{x} - \gamma x_i)^2}$$

And note that within an α group $\left(\widehat{\text{prem}}_p \alpha_i - \frac{\alpha_i}{\bar{\alpha}}\gamma\bar{x}\right)$ is constant. Allowing μ_α to be the age-weight for age group α,

$$1 - \frac{\sum_{i=1}^{N}\left(\widehat{\text{prem}}_p \alpha_i - \gamma x_i\right)^2}{\sum_{i=1}^{N}(\gamma\bar{x} - \gamma x_i)^2} = 1 - \frac{\sum_{i=1}^{N}\left[\left(\widehat{\text{prem}}_p \alpha_i - \frac{\alpha_i}{\bar{\alpha}}\gamma\bar{x}\right)^2 + \left(\frac{\alpha_i}{\bar{\alpha}}\gamma\bar{x} - \gamma x_i\right)^2\right] + 2 * \sum_{\alpha}\left(\widehat{\text{prem}}_p \mu_\alpha - \frac{\mu_\alpha}{\bar{\alpha}}\gamma\bar{x}\right)\sum_{i=1}^{N_\alpha}\left(\frac{\alpha_i}{\bar{\alpha}}\gamma\bar{x} - \gamma x_i\right)}{\sum_{i=1}^{N}(\gamma\bar{x} - \gamma x_i)^2}$$

Now, as long as the age curve is "close" to being accurate, i.e. $\frac{\mu_\alpha}{\bar{\alpha}}\gamma\bar{x} \approx \gamma\bar{x}_\alpha$, then the cross term goes to zero: $\sum_\alpha \left(\widehat{\mathrm{prem}}_p \mu_\alpha - \frac{\mu_\alpha}{\bar{\alpha}}\gamma\bar{x}\right)\sum_{i=1}^{N}\left(\frac{\alpha_i}{\bar{\alpha}}\gamma\bar{x} - \gamma x_i\right) \approx 0$ so that

$$1 - \frac{\sum_{i=1}^{N}\left(\widehat{\mathrm{prem}}_p \alpha_i - \gamma x_i\right)^2}{\sum_{i=1}^{N}(\gamma\bar{x} - \gamma x_i)^2} \approx 1 - \left[\frac{\sum_{i=1}^{N}\left(\frac{\alpha_i}{\bar{\alpha}}\gamma\bar{x} - \gamma x_i\right)^2}{\sum_{i=1}^{N}(\gamma\bar{x} - \gamma x_i)^2} + \frac{\sum_{i=1}^{N}\left(\widehat{\mathrm{prem}}_p \alpha_i - \frac{\alpha_i}{\bar{\alpha}}\gamma\bar{x}\right)^2}{\sum_{i=1}^{N}(\gamma\bar{x} - \gamma x_i)^2}\right] \tag{B1}$$

Note that as the age curve becomes less accurate, the approximation (and the measures derived from it) becomes worse.

Now, it is clear that the first component of the expression (premium fit) is the R-squared from a regression of x on the age weights, and all necessary pieces of data will be available in an *ex ante* claims dataset. The second component requires additional derivations.

First, we need to determine how premiums are set. As before, we assume that competition forces insurers to set base prices equal to average net costs. Then the base premium for the gold plan can be written

$$\widehat{\mathrm{prem}}_p^g = \frac{(\gamma + 1)\bar{x}_p^g}{\bar{\alpha}^g}$$

And for the silver plan

$$\widehat{\mathrm{prem}}_p^s = \frac{\bar{x}_p^s}{\bar{\alpha}^s}$$

Individual gold premiums are then equal to

$$\mathrm{prem}_{p,i}^g = \widehat{\mathrm{prem}}_p^g \alpha_i = (\gamma + 1)\bar{x}_p^g \frac{\alpha_i}{\bar{\alpha}^g}$$

And individual incremental premiums are equal to

$$\mathrm{prem}_{i,a}^e = \alpha_i\left(\frac{(\gamma + 1)\bar{x}_p^g}{\bar{\alpha}^g} - \frac{\bar{x}_p^s}{\bar{\alpha}^s}\right)$$

If we assume that $\bar{\alpha}^g = \bar{\alpha}^s = \bar{\alpha}$, then

$$\mathrm{prem}_{i,a}^e = \frac{\alpha_i}{\bar{\alpha}}\left((\gamma + 1)\bar{x}_p^g - \bar{x}_p^s\right)$$

Now, plugging this expression back into the second component of (B1), we get that

$$\frac{\sum_{i=1}^{N}\left(\widehat{\mathrm{prem}}_p \alpha_i - \frac{\alpha_i}{\bar{\alpha}}\gamma\bar{x}\right)^2}{\sum_{i=1}^{N}(\gamma\bar{x} - \gamma x_i)^2} = \frac{\sum_{i=1}^{N}\left(\frac{\alpha_i}{\bar{\alpha}}\left((\gamma + 1)\bar{x}_p^g - \bar{x}_p^s\right) - \frac{\alpha_i}{\bar{\alpha}}\gamma\bar{x}\right)^2}{\sum_{i=1}^{N}(\gamma\bar{x} - \gamma x_i)^2}$$

$$\frac{\sum_{i=1}^{N}\left(\widehat{\mathrm{prem}}_p \alpha_i - \frac{\alpha_i}{\bar{\alpha}}\gamma\bar{x}\right)^2}{\sum_{i=1}^{N}(\gamma\bar{x} - \gamma x_i)^2} = \frac{\sum_{i=1}^{N}\left(\frac{\alpha_i}{\bar{\alpha}}\left((\gamma + 1)\bar{x}_p^g - \bar{x}_p^s - \gamma\bar{x}\right)\right)^2}{\sum_{i=1}^{N}(\gamma\bar{x} - \gamma x_i)^2}$$

$$\frac{\sum_{i=1}^{N}\left(\widehat{prem}_p\alpha_i - \frac{\alpha_i}{\bar{\alpha}}\gamma\bar{x}\right)^2}{\sum_{i=1}^{N}(\gamma\bar{x}-\gamma x_i)^2} = \frac{\sum_{i=1}^{N}\left(\frac{\alpha_i}{\bar{\alpha}}\right)^2\left(\left((\gamma+1)\bar{x}_p^g - \bar{x}_p^s - \gamma\bar{x}\right)\right)^2}{\gamma^2\sum_{i=1}^{N}(x_i-\bar{x})^2}$$

By assumption 1 we know that

$$\left((\gamma+1)\bar{x}_p^g - \bar{x}_p^s - \gamma\bar{x}\right)^2 = \left(\gamma\left(\bar{x}_p^g - \bar{x}\right) + \left(\bar{x}_p^g - \bar{x}_p^s\right)\right)^2 = [\gamma^2\eta^2 + 1 + 2\gamma\eta]\left(\bar{x}_p^g - \bar{x}_p^s\right)^2$$

This allows us to simplify to get

$$\frac{\sum_{i=1}^{N}\left(\widehat{prem}_p\alpha_i - \frac{\alpha_i}{\bar{\alpha}}\gamma\bar{x}\right)^2}{\sum_{i=1}^{N}(\gamma\bar{x}-\gamma x_i)^2} = \left[\eta^2 + \frac{1}{\gamma^2} + \frac{2\eta}{\gamma}\right]\frac{\sum_{i=1}^{N}\left(\frac{\alpha_i}{\bar{\alpha}}\right)^2\left(\bar{x}_p^g - \bar{x}_p^s\right)^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2}$$

$$\frac{\sum_{i=1}^{N}\left(\widehat{prem}_p\alpha_i - \frac{\alpha_i}{\bar{\alpha}}\gamma\bar{x}\right)^2}{\sum_{i=1}^{N}(\gamma\bar{x}-\gamma x_i)^2} = \left[\eta^2 + \frac{1}{\gamma^2} + \frac{2\eta}{\gamma}\right]\frac{\left(\bar{x}_p^g - \bar{x}_p^s\right)^2\sum_{i=1}^{N}\left(\frac{\alpha_i}{\bar{\alpha}}\right)^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2}$$

which, by the derivations in Appendix A, can be written as

$$\frac{\sum_{i=1}^{N}\left(\widehat{prem}_p\alpha_i - \frac{\alpha_i}{\bar{\alpha}}\gamma\bar{x}\right)^2}{\sum_{i=1}^{N}(\gamma\bar{x}-\gamma x_i)^2} = \left[\eta^2 + \frac{1}{\gamma^2} + \frac{2\eta}{\gamma}\right]\left(1-(\delta^g+\delta^s)\right)\left(\frac{1}{\tau(1-\tau)}\right)\frac{\left[\sum_{i=1}^{N}\left(x_i - r_{p,i}\right)^2\right]\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\alpha_i}{\bar{\alpha}}\right)^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2}$$

$$\frac{\sum_{i=1}^{N}\left(\widehat{prem}_p\alpha_i - \frac{\alpha_i}{\bar{\alpha}}\gamma\bar{x}\right)^2}{\sum_{i=1}^{N}(\gamma\bar{x}-\gamma x_i)^2} = \left[\eta^2 + \frac{1}{\gamma^2} + \frac{2\eta}{\gamma}\right]\left(1-(\delta^g+\delta^s)\right)\left(\frac{1}{\tau(1-\tau)}\right)\frac{\left[\sum_{i=1}^{N}\left(x_i - r_{p,i}\right)^2\right]}{\sum_{i=1}^{N}(x_i-\bar{x})^2}\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\alpha_i}{\bar{\alpha}}\right)^2$$

$$\frac{\sum_{i=1}^{N}\left(\widehat{prem}_p\alpha_i - \frac{\alpha_i}{\bar{\alpha}}\gamma\bar{x}\right)^2}{\sum_{i=1}^{N}(\gamma\bar{x}-\gamma x_i)^2} = \omega\frac{\left[\sum_{i=1}^{N}\left(x_i - r_{p,i}\right)^2\right]}{\sum_{i=1}^{N}(x_i-\bar{x})^2}\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\alpha_i}{\bar{\alpha}}\right)^2$$

This allows us to write $\tilde{\phi}^{age}$ as a function of $\tilde{\phi}$:

$$\tilde{\phi}^{age} = \tilde{\phi}\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\alpha_i}{\bar{\alpha}}\right)^2$$

And, analogously, the final measure with the age-curve can be written as a function of the normal measure under the single premium system:

$$\text{Payment System Fit}^{age} = 1 - \phi^{age} = 1 - \frac{\tilde{\phi}}{\omega}\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\alpha_i}{\bar{\alpha}}\right)^2$$

The intuition behind this measure is that there are two things that affect an individual's price with an age curve. First, because under the age-curve payment system everyone is effectively in the same risk pool, the variance of net costs across all individuals helps determine the base incremental premium. Second, the age curve causes different age groups to pay different premiums while being in the same risk pool. The age curve effectively magnifies the incremental premium for the old and shrinks it for the young. This presents the need for the $\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\alpha_i}{\bar{\alpha}}\right)^2$ term in the payment system fit measure.

Appendix C: Matching and Selection Methods

In Section 5.1 of the main text, we describe the sample of individuals whose health insurance claims data we use to estimate our selection measures. As discussed in that section, we use a sample of individuals from the 2012-2013 Truven Analytics Marketscan Database selected to mimic the group of individuals eligible to enroll in the Health Insurance Marketplaces. We choose our sample by identifying a group of Marketplace-eligible individuals in the MEPS and by using propensity score methods to determine which individuals in the MarketScan Database look most similar to the MEPS sample. Here, we explain this process in detail.

The Medical Expenditure Panel Survey (MEPS) is a large, nationally representative survey of the civilian non-institutionalized U.S. population with information on approximately 33,000 individuals annually. We pool MEPS data from Panels 9 (2004/5) through 16 (2011/12) and select a population of individuals we designate as Marketplace eligible based on income, insurance, Medicaid eligibility, and employment status using the requirements outlined in the ACA.[63] Each person in MEPS is present for two years in the data. We apply our inclusion and exclusion criteria annually, and include an individual if they satisfy the inclusion criteria in either year. We include adult, non-elderly individuals (19-64 in the first year of the data) in households earning at least 138 percent of the federal poverty level (FPL) and children in households with income of at least 205 percent of FPL. Selection criteria into the Marketplace population, as defined by the ACA, include individuals from these groups who are, or in the case of children those who live in, a household where an adult is: ever uninsured, a holder of a non-group insurance policy, or a holder of a self-employed insurance policy. We ignore the premium affordability criteria, as we focus on the individual Marketplaces and those who pay high group premiums would probably enter the special small group market (SHOP) or remain in group coverage.[64] We also exclude individuals continuously enrolled in Medicaid for the entire year. Marketplaces apply separate risk adjustment formula to children and adults. The adult risk adjustment formula is applied to Marketplace enrollees 21-64. In this analysis, we focus on adults 21-64, not children.

In order to prepare for application of a matching algorithm to the MarketScan data, we identify a second group in MEPS, those adults who have coverage similar to that in the MarketScan sample. Specifically, we identify the group of individuals who at any time during the two-year survey period have employer-sponsored health insurance and work in a firm with more than 100 employees. We call this set of people the Large Group sample. Our sample selection criteria when applied to MEPS identified 17,377 individuals with a designation of Marketplace Eligible and 19,964 with a designation of Large Group.

The Truven Analytics MarketScan data is a large longitudinal enrollment and claims databases from large employers and large insurers in the U.S. MarketScan data from 2010 were used

---

[63] We used a similar data selection process in MEPS in McGuire et al (2013) and McGuire et al. (2014). Here we add one more MEPS panel.

[64] The premium affordability criterion is that individuals would also be eligible for the Marketplace if they are covered by group insurance with premiums more than 9.5% of their income. SHOP refers to the Small Business Health Options Program, part of the ACA.

to calibrate the HHS-HCC risk adjustment system for Marketplaces applied beginning in 2014, and are likely candidates for future recalibration. Here, we use MarketScan data from 2012 and 2013, applying the same criteria used by the developers of the present Marketplace risk adjustment formula.

Following criteria applied in the HHS-HCC model developed for the Marketplaces for 2014 (Kautter et al, 2014), we keep observations for individuals who are continuously eligible for coverage, enrolled in a preferred provider organization or other fee-for-service (FFS) health plan in both the first and last month of a year,[65] and have no payments made on a capitated basis. Also, following HHS criteria, we required individuals to have both mental health and drug coverage. We also exclude individuals who have claims with negative payments. Since we are estimating both prospective and concurrent risk adjustment models, we require individuals continuously eligible for FFS coverage in both 2012 and 2013. Altogether 7,072,964 individuals in MarketScan meet our inclusion and exclusion criteria.

From this Marketscan sample of potentially matchable eligible adults we seek a subset of the population with characteristics similar to adults eligible for Marketplace coverage. Our approach is to identify a set of variables available in both the MEPS and Marketscan samples and then use propensity score matching methods to identify a sample of individuals in the Marketscan sample that matches the joint distribution of the matching variables in the MEPS sample. The matching variables we use are: age, gender, region, an indicator for residence in a Metropolitan Statistical Area (MSA), an indicator for an inpatient admission in each of ten diagnostic categories (including a large "other" category), the total number of inpatient admissions, and quantile of outpatient and prescription drug spending. All variables are defined from claims and eligibility data from 2012 for MarketScan and the first year of each MEPS two-year panel, except for the diagnostic indicators for admissions and the count of admissions, for which we use 2012 *and* 2013 data for MarketScan and both panel years for MEPS.

To categorize admissions in MarketScan, we use the ICD-9 principle diagnosis. These are aggregated to Clinical Classification Software (CCS) codes, and are further classified into 10 broader categories as described in Table C1, paralleling the earlier study (McGuire et al., 2014). MEPS has a CCS indicator for each admission. To define spending distributions, we specify unequal quantiles, in order to capture the skewness in spending distributions. Specifically, for both outpatient and prescription drug spending we define four unequal quantiles: 0-50%, 50-80%, 80-90%, and 90-100%. The variables in the matching file are summarized in Table C2. Income has been inflated to 2012 dollars according to CPI-U, and health spending has been inflated to 2012 dollars according to CPI-Medical Care.

The propensity score model was built in the MEPS data using an indicator for being in the Marketplace-eligible group as the outcome. All variables except the person ID, weight variables, age, and sex were used to fit a weighted regression, where the weight was the MEPS individual weight used to make the MEPS data nationally representative, perwt99f_y1. Age and sex were not

[65] Other plan types included are Exclusive Provider Organization, Non-Capitated Point-of-Service, Consumer-Driven Health Plan, and High-Deductible Health Plan.

included in favor of the age/sex category bins as described in Table C2. An indicator was also dropped for each set of categorical variables.

We drew a matched sample of subjects from MarketScan that had similar characteristics to those who were Exchange eligible in MEPS. This involved the use of common variables to select individuals in MarketScan such that the distribution of these common variables was similar in MEPS and MarketScan. Each subject in MarketScan received a predicted value (propensity score) based on the propensity score model estimated among MEPS subjects. We matched K= perwt99f_y1/64 MarketScan subjects to each Exchange-eligible MEPS subject based on these propensity score values, where K is a scaled nationally-representative weight given to each MEPS observation. The nationally representative weight was divided by 64 to yield a sample size of Marketscan subjects that was at least 2 million. The match involved finding subjects in MarketScan who had similar propensity score values, where the greatest distance allowed was defined by c=SD(logit(propensity score))/4.

Appendix D: Risk Adjustment Estimation

As discussed in the text, we simulate insurer revenues under payment systems incorporating both concurrent and prospective risk adjustment. In order to simulate revenues under risk adjustment, we first have to assign risk scores to each individual. Risk scores are assigned using the following formula:

$$\text{risk}_{pi} = \beta z_i \tag{D1}$$

where risk_{pi} is person i's risk score under payment system p, z_i is a vector of risk adjustment variables, and β is a vector of weights describing the contribution of each risk adjustment variable to the risk score. Each of these variables differs for the concurrent and prospective models. In addition to the concurrent and prospective risk scores, we also assign each individual a risk score based on a simple age/gender risk adjustment model. Denote the variables for model j as risk^j_{pi} , z^j_i, and β^j where $j \in \{\text{conc}, \text{pros}, \text{age}\}$.

For the concurrent and prospective risk adjustment models, z^j_i consists of the same variables included in the HHS-HCC risk adjustment model. The model includes a set of 18 mutually exclusive age-by-sex indicators, 91 indicators for chronic conditions referred to as Hierarchical Condition Categories, and 19 interaction terms capturing interactions between different sets of chronic conditions.[66] We assign each z^j_i variable to each individual using publicly available software provided by HHS.[67] The software takes diagnoses and information on the individual's age and sex and outputs each of the variables in the risk adjustment model. For the prospective model we provide the software with diagnoses from 2012, and for the concurrent model we provide the software with diagnoses from 2013. This results in z^{pros}_i and z^{conc}_i. For the age/gender model, we use the age/gender categories output by the software when generating the prospective (2012 diagnoses) variables, producing z^{age}_i.

We estimate the weights for each model, β^j, using the same method employed by the policymakers who developed the HHS-HCC model. While we could have used the weights estimated by the policymakers, they only estimated the concurrent weights, so to make sure that the prospective, concurrent, and age/gender models were on a "level playing field" we estimate all of the weights ourselves. We estimate the weights via a linear regression of *ex post* annual spending, x_i, on z^j_i using our 2013 MarketScan Marketplace (MM) sample.[68] To protect against "overfitting," we estimate the weights using a k-fold cross validation procedure. To do this, we randomly divide all of the individuals in the MM sample into 10 groups. For the first group, we estimate the weights, β^j, via a linear regression of x_i on z^j_i using data from the other 9 groups. We then repeat the same procedure for the other 9 groups. This ensures that the weights for each group are not estimated using data from individuals in that group. The estimated weights for each model (averaged across the 10 groups) are found in Table D1.

Raw risk scores for the concurrent risk adjustment model were assigned according to Equation D1. For the prospective model, we randomly divided the MM sample into 2 groups: incumbent enrollees and new enrollees. The new enrollees were assigned the risk score implied by

[66] The 91 categories consist of 74 HCCs and 17 groups of other HCCs not already included in the model.
[67] The software can be found here: http://www.cms.gov/CCIIO/Resources/Regulations-and-Guidance/Downloads/SASsoftware.zip
[68] Note that because our MM sample only includes individuals with 24 continuous months of coverage through 2012 and 2013, there was no need to "annualize" spending as done in Kautter et al. (2014). Also note that we regress total spending (plan paid + out-of-pocket) on z^j_i rather than just the portion paid by the plan.

the age/gender risk adjustment model, and the incumbent enrollees were assigned the score implied by the prospective risk adjustment model:

$$\text{risk}_{pi}^{pros} = \begin{cases} \beta^{age} x_i^{age} & \text{if new} \\ \beta^{pros} x_i^{pros} & \text{if incumbent} \end{cases}$$

Relative risk scores were calculated by dividing each individual's raw risk score by the average raw risk score in the population.

Appendix E: Prediction Models for Expected Spending

In Section 6.3, we test the robustness of two of our measures, premium fit and the measure developed in Section 4, to various assumptions about the specification of expected spending. As discussed in that section, we assign each individual a "predicted cost" and then re-estimate our measures replacing *ex post* spending with a weighted average of the predicted cost and *ex post* spending, varying the weight on *ex post* spending from 0% to 100%. Here, we discuss how we estimate each individual's predicted cost.

We specify an individual's 2013 predicted cost to be a linear function of the same age/gender cells used in the HHS-HCC risk adjustment model plus indicators for the individual's decile of 2012 inpatient spending, outpatient spending, and prescription drug spending. More formally, let AgeSex_i^j be a dummy variable equal to 1 if person i falls into age/gender cell j. Similarly, let IP_i^j, OP_i^j, and RX_i^j be dummy variables equal to 1 if person i's 2012 spending put them in decile j of 2012 inpatient, outpatient, or prescription drug spending in the population, respectively. We define deciles in an atypical way, reflecting the large portion of individuals who have zero spending in each category. We define the first decile as all individuals with zero spending in that category, and then we divide the remaining individuals who have positive spending into 9 equally sized groups.

We specify person i's 2013 predicted spending to be equal to the predicted value from the following regression:

$$x_i = \sum \zeta^j \mathrm{AgeSex}_i^j + \sum \kappa^j \mathrm{IP}_i^j + \sum \lambda^j \mathrm{OP}_i^j + \sum \tau^j \mathrm{RX}_i^j$$

The estimated values for ζ^j, κ^j, λ^j, and τ^j are found in Table E1. We repeat this exercise to generate predicted inpatient, outpatient, and prescription drug spending, replacing x_i with 2012 inpatient, outpatient, and prescription drug spending, respectively, while keeping the right-hand side of the equation the same.

Table C1. 10 Groups of CCS codes

1 Heart Disease (96, 97, 100-108)
2 Injury (225-236, 239, 240, 244)
3 Cancer (11-45)
4 Mental Health and Substance Abuse (650-663)
5 Lower Respiratory Disorders (122, 127-133)
6 Diabetes (49, 50)
7 Non-Traumatic Joint and Back Disorders (201-205)
8 Pregnancy and/or Delivery (181-196)
9 Kidney Diseases (156-158, 160, 161)
10 All Other
* Numbers in the parenthesis are CCS codes included in each group.

Table C2. Variables in MEPS File

Category	Variable Name	Description
Identifier	Dupersid	person id
	Core	Exchange eligibility, core=1 if Exchange eligible
Demographics		
	Age	first year age
	Sex	first year gender, 1=male, 2=female
	age-gender categories	a group of dummy variables, 5 year bands for people with age 21-54, and 2 year bands for people with age 55-64
	Msa	Metropolitan Statistical Area, 1=live in MSA in first-year
	region_northeast	1=live in northeast in the first year
	region_midwest	1=live in midwest in the first year
	region_south	1=live in south in the first year
	region_west	1=live in west in the first year
Inpatient diagnosis		
	ip_ccs1_2yrs	1= has the inpatient diagnosis in CCS1 in either year
	ip_ccs2_2yrs	1= has the inpatient diagnosis in CCS2 in either year
	ip_ccs3_2yrs	1= has the inpatient diagnosis in CCS3 in either year
	ip_ccs4_2yrs	1= has the inpatient diagnosis in CCS4 in either year
	ip_ccs5_2yrs	1= has the inpatient diagnosis in CCS5 in either year
	ip_ccs6_2yrs	1= has the inpatient diagnosis in CCS6 in either year
	ip_ccs7_2yrs	1= has the inpatient diagnosis in CCS7 in either year
	ip_ccs8_2yrs	1= has the inpatient diagnosis in CCS8 in either year
	ip_ccs9_2yrs	1= has the inpatient diagnosis in CCS9 in either year
	ip_ccs10_2yrs	1= has the inpatient diagnosis in CCS10 in either year
	ip_adm_0	1= no inpatient admissions in two-year periods
	ip_adm_1	1= 1 inpatient admissions in two-year periods
	ip_adm_2	1= 2 inpatient admissions in two-year periods
	ip_adm_3	1= 3 inpatient admissions in two-year periods
	ip_adm_4p	1= 4 or more inpatient admissions in two-year periods
Health Spending		
	opexp_below50	1= outpatient spending below 50% quantile in first year
	opexp_5080	1= outpatient spending between 50%-80% in first year
	opexp_8090	1= outpatient spending between 80-90% in first year
	opexp_90above	1= outpatient spending above 90% in first year
	rxexp_below50	1= drug spending below 50% quantile in first year
	rxexp_5080	1= drug spending between 50%-80% in first year
	rxexp_8090	1= drug spending between 80-90% in first year
	rxexp_90above	1= drug spending above 90% in first year
Other	perwt99f_y1	weight of the person in the first year
	perwt99f_y2	weight of the person in the second year

Table D1: Estimated Risk Adjustment Model Weights

Variable	Prospective	Concurrent	Age/Gender
Intercept	1.168	0.751	1.893
Male 21-24	-0.973	-0.592	-1.589
Male 25-29	-0.944	-0.577	-1.572
Male 30-34	-0.865	-0.527	-1.449
Male 35-39	-0.772	-0.473	-1.313
Male 40-44	-0.693	-0.415	-1.187
Male 45-49	-0.538	-0.335	-0.946
Male 50-54	-0.334	-0.244	-0.623
Male 55-59	-0.079	-0.167	-0.209
Male > 60	0.101	-0.028	0.163
Female 21-24	-0.715	-0.456	-1.296
Female 25-29	-0.406	-0.368	-0.950
Female 30-34	-0.404	-0.288	-0.917
Female 35-39	-0.468	-0.249	-0.937
Female 40-44	-0.469	-0.203	-0.885
Female 45-49	-0.363	-0.161	-0.698
Female 50-54	-0.250	-0.098	-0.489
Female 55-59	-0.158	-0.084	-0.281
Female > 60	omitted	omitted	omitted
HIV/AIDS	4.327	4.128	
Septicemia, Sepsis, Systemic Inflammatory Response Syndrome/Shock	2.497	10.118	
Central Nervous System Infections, Except Viral Meningitis	1.396	4.238	
Viral or Unspecified Meningitis	-0.415	2.379	
Opportunistic Infections	2.234	4.578	
Metastatic Cancer	13.989	16.375	
Lung, Brain, and Other Severe Cancers, Including Pediatric Acute Lymphoid Leukemia	8.922	8.544	
Non-Hodgkin`s Lymphomas and Other Cancers and Tumors	3.332	4.068	
Colorectal, Breast (Age < 50), Kidney, and Other Cancers	2.543	4.033	
Breast (Age 50+) and Prostate Cancer, Benign/Uncertain Brain Tumors, and Other Cancers and Tumors	1.392	2.311	
Thyroid Cancer, Melanoma, Neurofibromatosis, and Other Cancers and Tumors	0.817	1.250	
Pancreas Transplant Status/Complications	6.924	6.112	
Protein-Calorie Malnutrition	4.877	9.653	

Liver Transplant Status/Complications	7.538	20.413
End-Stage Liver Disease	12.056	4.073
Cirrhosis of Liver	3.772	1.593
Chronic Hepatitis	1.178	1.670
Acute Liver Failure/Disease, Including Neonatal Hepatitis	0.918	2.525
Intestine Transplant Status/Complications	0.693	10.423
Peritonitis/Gastrointestinal Perforation/Necrotizing Enterocolitis	1.960	14.033
Intestinal Obstruction	1.980	4.562
Chronic Pancreatitis	4.629	4.934
Acute Pancreatitis/Other Pancreatic Disorders and Intestinal Malabsorption	1.243	2.309
Inflammatory Bowel Disease	2.559	2.394
Rheumatoid Arthritis and Specified Autoimmune Disorders	3.005	2.568
Systemic Lupus Erythematosus and Other Autoimmune Disorders	1.280	0.805
Cleft Lip/Cleft Palate	1.366	1.249
Hemophilia	35.480	29.502
Coagulation Defects and Other Specified Hematological Disorders	1.175	2.135
Schizophrenia	2.012	2.566
Major Depressive and Bipolar Disorders	1.216	1.342
Reactive and Unspecified Psychosis, Delusional Disorders	1.336	3.172
Personality Disorders	1.217	1.363
Anorexia/Bulimia Nervosa	1.412	2.230
Prader-Willi, Patau, Edwards, and Autosomal Deletion Syndromes	2.188	4.704
Down Syndrome, Fragile X, Other Chromosomal Anomalies, and Congenital Malformation Syndromes	0.524	-0.027
Autistic Disorder	0.432	0.167
Pervasive Developmental Disorders, Except Autistic Disorder	0.416	0.641
Spinal Cord Disorders/Injuries	1.172	5.449
Amyotrophic Lateral Sclerosis and Other Anterior Horn Cell Disease	5.580	4.826
Quadriplegic Cerebral Palsy	13.157	6.415
Cerebral Palsy, Except Quadriplegic	0.543	0.197
Spina Bifida and Other Brain/Spinal/Nervous System Congenital Anomalies	1.954	-0.428

Myasthenia Gravis/Myoneural Disorders and Guillain-Barre Syndrome/Inflammatory and Toxic Neuropathy	2.973	4.187
Multiple Sclerosis	7.549	7.111
Seizure Disorders and Convulsions	1.645	6.366
Hydrocephalus	1.690	7.025
Non-Traumatic Coma, Brain Compression/Anoxic Damage	2.351	9.391
Respirator Dependence/Tracheostomy Status	7.092	27.955
Congestive Heart Failure	2.439	2.606
Acute Myocardial Infarction	1.688	7.322
Unstable Angina and Other Acute Ischemic Heart Disease	1.121	4.095
Heart Infection/Inflammation, Except Rheumatic	0.457	3.802
Specified Heart Arrhythmias	1.252	2.380
Intracranial Hemorrhage	1.463	6.753
Ischemic or Unspecified Stroke	1.125	2.766
Cerebral Aneurysm and Arteriovenous Malformation	1.790	4.026
Hemiplegia/Hemiparesis	1.419	5.021
Monoplegia, Other Paralytic Syndromes	1.213	3.189
Atherosclerosis of the Extremities with Ulceration or Gangrene	5.219	7.635
Vascular Disease with Complications	2.108	5.210
Pulmonary Embolism and Deep Vein Thrombosis	1.556	7.699
Lung Transplant Status/Complications	20.605	43.748
Cystic Fibrosis	11.712	9.934
Fibrosis of Lung and Other Lung Disorders	1.719	1.605
Aspiration and Specified Bacterial Pneumonias and Other Severe Lung Infections	1.656	4.496
Kidney Transplant Status	5.451	6.380
End Stage Renal Disease	28.792	24.537
Chronic Ulcer of Skin, Except Pressure	2.572	1.807
Hip Fractures and Pathological Vertebral or Humerus Fractures	2.146	6.307
Pathological Fractures, Except of Vertebrae, Hip, or Humerus	0.835	1.567
Stem Cell, Including Bone Marrow, Transplant Status/Complications	7.515	19.868
Artificial Openings for Feeding or Elimination	2.156	5.743
Amputation Status, Lower Limb/Amputation Complications	1.833	4.036
Group 1	1.379	1.038
Group 2	1.472	1.443

Group 3	2.451	4.953
Group 4	2.233	2.552
Group 6	13.044	10.437
Group 7	4.243	4.634
Group 8	3.497	3.156
Group 9	2.455	3.045
Group 10	6.401	10.766
Group 11	4.258	7.094
Group 12	2.618	1.596
Group 13	2.550	11.649
Group 14	12.082	30.263
Group 15	0.912	0.822
Group 16	5.234	1.178
Group 17	1.270	0.894
Group 18	-0.054	2.633
Interaction Group M	0.698	-3.284
Interaction Group H	0.729	-15.242
Severe V3	-0.428	-5.480
SEVERE_V3_x_HHS_HCC006	-1.105	22.307
SEVERE_V3_x_HHS_HCC008	2.582	18.013
SEVERE_V3_x_HHS_HCC009	-2.181	17.873
SEVERE_V3_x_HHS_HCC010	1.752	19.153
SEVERE_V3_x_HHS_HCC115	-1.812	11.488
SEVERE_V3_x_HHS_HCC135	1.359	20.228
SEVERE_V3_x_HHS_HCC145	-1.882	17.204
SEVERE_V3_x_G06	-13.682	10.319
SEVERE_V3_x_G08	10.617	19.391
SEVERE_V3_x_HHS_HCC035	-10.007	6.682
SEVERE_V3_x_HHS_HCC038	-1.236	2.284
SEVERE_V3_x_HHS_HCC153	-5.542	-0.929
SEVERE_V3_x_HHS_HCC154	0.356	4.570
SEVERE_V3_x_HHS_HCC163	-0.623	3.415
SEVERE_V3_x_HHS_HCC253	-2.487	8.126
SEVERE_V3_x_G03	0.510	9.427

Notes: The table reports the (average) normalized estimated coefficients for each risk adjustment variable for the prospective (prior year diagnoses), concurrent (current year diagnoses), and age/gender models. Coefficients are estimated using k-fold techniques described in the text to protect against “overfitting.” The coefficients reported in the table are the average of the estimated coefficients across the k groups. Coefficients are normalized by dividing by the average cost in the population, so that a risk score of 2.0 implies the individual is expected to have costs equal to 2 times the average cost. The coefficients are estimated on the Marketscan Marketplace sample.

Table D2: Federal Age Curve

Age	HHS Age Curve Weights	Age	HHS Age Curve Weights
21	1	43	1.357
22	1	44	1.397
23	1	45	1.444
24	1	46	1.5
25	1.004	47	1.563
26	1.024	48	1.635
27	1.048	49	1.706
28	1.087	50	1.786
29	1.119	51	1.865
30	1.135	52	1.952
31	1.159	53	2.04
32	1.183	54	2.135
33	1.198	55	2.23
34	1.214	56	2.333
35	1.222	57	2.437
36	1.23	58	2.548
37	1.238	59	2.603
38	1.246	60	2.714
39	1.262	61	2.81
40	1.278	62	2.873
41	1.302	63	2.952
42	1.325	64	3

Table E1: Estimated Risk Adjustment Model Weights

Variable	Total	IP	OP	RX
Intercept	59809.98	25294.33	25505.81	9009.84
Male 21-24	-2357.07	-947.43	-1486.17	76.54
Male 25-29	-2249.62	-937.28	-1430.86	118.52
Male 30-34	-2181.57	-913.44	-1396.04	127.91
Male 35-39	-1930.17	-801.77	-1283.90	155.50
Male 40-44	-1676.23	-642.00	-1185.08	150.85
Male 45-49	-1233.19	-417.94	-973.84	158.58
Male 50-54	-598.11	14.59	-730.04	117.34
Male 55-59	383.19	650.33	-368.45	101.31
Male > 60	1115.75	939.37	111.08	65.31
Female 21-24	-2105.67	-770.53	-1373.78	38.64
Female 25-29	-1292.77	-242.64	-1112.41	62.27
Female 30-34	-1227.73	-378.12	-946.54	96.93
Female 35-39	-1337.50	-701.10	-752.11	115.71
Female 40-44	-1404.13	-845.10	-675.91	116.89
Female 45-49	-1044.56	-733.26	-424.47	113.17
Female 50-54	-830.54	-558.00	-380.53	107.99
Female 55-59	-564.80	-355.27	-285.10	75.57
Female > 60	omitted	omitted	omitted	omitted
Outpatient Decile 1	-10313.63	-2888.16	-6701.02	-724.45
Outpatient Decile 2	-10144.61	-2913.07	-6502.46	-729.09
Outpatient Decile 3	-10096.14	-2948.70	-6421.25	-726.19
Outpatient Decile 4	-9946.72	-2910.46	-6313.63	-722.63
Outpatient Decile 5	-9806.36	-2910.45	-6168.80	-727.11
Outpatient Decile 6	-9546.67	-2850.92	-5991.42	-704.32
Outpatient Decile 7	-8931.24	-2604.63	-5648.41	-678.20
Outpatient Decile 8	-8199.03	-2400.75	-5197.29	-601.00
Outpatient Decile 9	-7105.40	-2061.19	-4518.05	-526.16
Outpatient Decile 10	omitted	omitted	omitted	omitted
Inpatient Decile 1	-33800.03	-18800.13	-13630.51	-1369.39
Inpatient Decile 2	-31651.99	-17533.77	-12791.47	-1326.76
Inpatient Decile 3	-31565.50	-17414.97	-12953.71	-1196.82
Inpatient Decile 4	-31332.28	-17473.79	-12649.28	-1209.22
Inpatient Decile 5	-30281.66	-16791.58	-12344.84	-1145.24
Inpatient Decile 6	-30012.53	-16830.76	-12038.53	-1143.23
Inpatient Decile 7	-27802.92	-15745.36	-10948.43	-1109.13
Inpatient Decile 8	-25727.99	-14668.83	-10016.48	-1042.68
Inpatient Decile 9	-20274.86	-11901.62	-7562.81	-810.44
Inpatient Decile 10	omitted	omitted	omitted	omitted
Prescription Drug Decile 1	-12853.38	-2597.95	-3294.55	-6960.89
Prescription Drug Decile 2	-12965.08	-2648.13	-3363.68	-6953.27

Prescription Drug Decile 3	-12934.92	-2664.84	-3330.95	-6939.13
Prescription Drug Decile 4	-12782.24	-2621.85	-3256.70	-6903.69
Prescription Drug Decile 5	-12670.91	-2601.90	-3205.91	-6863.10
Prescription Drug Decile 6	-12323.83	-2493.62	-3054.17	-6776.05
Prescription Drug Decile 7	-11884.46	-2448.36	-2860.23	-6575.87
Prescription Drug Decile 8	-11270.02	-2296.51	-2714.10	-6259.41
Prescription Drug Decile 9	-9878.78	-2035.35	-2334.46	-5508.96
Prescription Drug Decile 10	omitted	omitted	omitted	omitted

Notes: The table presents the estimated coefficients from a regression of total or service-level 2013 spending on the variables listed. The model is estimated on the Marketscan Marketplace sample, and it is used to construct a measure of each individual's 2013 predicted level of spending from the consumer's point of view.