Measuring efficiency of health plan payment systems in managed competition health insurance markets

Timothy J. Layton [a],[*], Randall P. Ellis [b], Thomas G. McGuire [c], Richard van Kleef [d]

[a] *Department of Health Care Policy, Harvard Medical School and NBER, United States*
[b] *Department of Economics, Boston University, United States*
[c] *Department of Health Care Policy, Harvard Medical School and NBER, United States*
[d] *Institute of Health Policy and Management, Erasmus University Rotterdam, The Netherlands*

A B S T R A C T

Adverse selection in health insurance markets leads to two types of inefficiency. On the demand side, adverse selection leads to plan price distortions resulting in 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. Reinsurance, risk adjustment, and premium categories address these problems. Building on prior research on health plan payment system evaluation, we develop measures of the efficiency consequences of price and benefit distortions under a given payment system. Our measures are based on explicit economic models of insurer behavior under adverse selection, incorporate multiple features of plan payment systems, and can be calculated prior to observing actual insurer and consumer behavior. We illustrate the use of these measures with data from a simulated market for individual health insurance.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

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. This natural pattern of demand causes two forms of economic inefficiency: 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 unprofitable (typically sicker) enrollees (Glazer and McGuire, 2000; Geruso and Layton, 2017). The relative importance of these two inefficiencies varies across markets. In the U.S. Medicare program, sorting beneficiaries between the private managed care plans (Medicare Advantage plans) and traditional Medicare is the main efficiency and policy issue (Curto et al., 2014), whereas in the national health insurance system in the Netherlands with common regulation and coverage for the entire population, underprovision of some services (*e.g.* exclusion of high-quality doctors or health care facilities from provider networks) is the main concern (van Kleef et al., 2013). Other markets, such as the Marketplaces

established in the U.S. as part of the Affordable Care Act (ACA) feature both concerns: inducing participation among those eligible to purchase coverage on the Marketplace (Newhouse, 2017) and ensuring that the plans provide adequate coverage for all conditions (Geruso et al., 2016). Economic analysis contends with both forms of selection-related inefficiencies by *ex ante* study of the incentives embedded in insurance markets under alternative policy environments[1] and by *ex post* evaluation of the performance of implemented policies based on actual consumer and insurer behavior.[2]

This paper develops and implements a general methodology for assessing the *ex ante* selection-related inefficiencies created by a health plan payment system in regulated individual health insurance markets. 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

---

[1] 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).

[2] See Ericson 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.

premium structure and the plan payment scheme – will contend with the two selection-related inefficiencies. Although *ex post* evaluation is obviously necessary and important, *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.[3] We believe the methodology of *ex ante* evaluation can be improved over the current state of the art. For many years, studies in this literature have focused on 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 include ratio or difference measures of over/under compensation. In the U.S., "predictive ratios," the ratio of predicted costs to actual costs for selected groups in the population, such as those with a chronic illness, are used, whereas in Europe researchers study over- and undercompensation – the difference between projected revenues and costs rather than their ratio. Although widely used, neither an R-squared nor a predictive ratio has been shown to have a direct interpretation in terms of economic efficiency.

In this paper we derive measures for *ex ante* evaluation of payment system performance in terms of economic efficiency, 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. Specifically, they should be based on the effects of a payment system on consumer welfare rather than just the (individual- or group-level) discrepancy between plan revenues and expected costs. 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.

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 ratio or difference measures of over/undercompensation. Other more economically valid welfare metrics, require estimates of key behavioral parameters among the population of interest, causing them to be impractical for *ex ante* policy analysis.[4] In a nutshell, our paper intends to equip researchers and regulators with a methodology, or a "toolkit," for evaluating payment system alternatives in terms of economic efficiency rather than statistical fit which has dominated this literature for more than 30 years.

Our focus is on 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 two measures, which we call "premium fit" and "payment system fit," are needed to 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. We show that premium fit is important independent of payment system fit due to the insight that even if a payment system perfectly matches revenues to costs at an individual level, inefficiency will remain because, except in a very special case, no single premium sorts consumers efficiently among plans. This links our analysis to papers by Bundorf et al. (2012), Geruso (2017), and others.

With respect to the second selection-related problem, existing service-by-service incentive measures based on predictability and predictiveness of each service do not produce an overall metric for efficiency (Ellis and McGuire, 2007). We generalize the service-by-service approach, and present a new cardinal measure that summarizes overall welfare loss developed in Layton et al. (2016a,b). This new metric can also be computed using the large administrative health insurance claims databases typically used for *ex ante* payment system evaluation.

We illustrate the use of our measures for the evaluation of the performance of the payment system being implemented in the ACA Marketplaces starting in 2017 (ACA 2017) relative to an alternative system. This policy remains a relevant baseline even if U.S. health policy moves away from this approach. Under ACA 2017, premiums are set by a pre-specified age curve, risk adjustment is concurrent (*i.e.* based on diagnoses from the current year rather than the prior year) and mandatory federal reinsurance is eliminated. We evaluate the plan against an alternative which keeps the existing age curve and reinstates reinsurance, but that switches to prospective risk adjustment (*i.e.* based on diagnoses from the prior year) rather than concurrent risk adjustment. Our comparison illustrates how our selection metrics work, and also provides relevant evidence for policy choices going forward for the Marketplaces.

We acknowledge that regulators have objectives in addition to reducing inefficiencies from adverse selection, including incentives for cost control, avoiding gameability, and fairness. Among the papers that consider these issues are Handel et al. (2015a) who study the tradeoff between inefficient plan choice and "reclassification risk," Beck et al. (2014) who consider the tradeoff between fairness and selection incentives, and Geruso and McGuire (2016) who evaluate a tradeoff between selection inefficiencies and incentives for cost control. We return in a final section to discuss other dimensions of comprehensive *ex ante* plan payment evaluation.

The remainder of the paper proceeds as follows: Section 2 reviews and critiques existing methods for assessing payment systems and measuring selection inefficiency and flags 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 health plan design (Section 4). We illustrate how our measures work in Section 5. We summarize and discuss the findings in Section 6.

---

[3] See as examples, Kautter et al. (2014) on US Marketplaces; Pope et al. (2011) on US Medicare; Shmueli et al. (2010) on Israel; Beck et al. (2010) on Switzerland; Breyer et al. (2003) on Germany; van Kleef et al. (2013) on the Netherlands.

[4] For example, the certainty-equivalent measure used by Einav et al. (2010a,b) requires estimates of risk aversion which are difficult to estimate from health insurance choices.

2. Assessing adverse selection in health insurance markets

The literature assessing inefficiencies related to 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 the Marketplaces as representative 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.[5] 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, the same data source used later in this paper, and separate models were estimated for different age categories and plan metal levels.[6] As of 2017, all states have elected to use the HHS-HCC model in their Marketplaces.

After developing the HHS-HCC model, Kautter and colleagues conducted an *ex ante* analysis of the Marketplace payment system, reporting two types of statistics: 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 14 m 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). 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.[7]

The statistic most commonly 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.

Other papers and reports compare model predictions and actual costs for subgroups defined in various ways. In their evaluation of the CMS-HCC model, Pope et al. (2011) 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.[8] 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.

In practice there are often 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 health plan incentives at both the individual and group level. 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. However, most of the risk adjustment literature reporting measures of statistical fit ignores these other payment system features.

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. (2016a) 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.[9] McGuire et al. (2014) and Geruso et al. (2016) modify predictive ratios in the same way. 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.

More fundamental is the missing underlying economic rationale for these statistical measures (even the more comprehensive "payment system measures"). While it is intuitive that better fit in both forms should improve the performance of a payment system with respect to selection problems, the interpretation in terms of economic efficiency remains unclear. One line of critique of R-squared measures argues that "only predictable costs matter" in assessing alternative risk adjustment models. However, it is not entirely clear why this is the case. Additionally, "predictability" is not a simple matter to incorporate into metrics of *ex ante* payment system performance.[10] One of the virtues of our approach of deriving met-

---

[5] 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 et al. (2012) discuss the attributes of good risk adjuster variables, and update the use of risk adjustment in health care systems internationally.

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

[7] 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 fit of the risk adjustment formula. 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).

[8] The "mean prediction error" for subgroups, as figured in Hsu et al. (2010), is similar.

[9] 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 $\frac{\text{Var}\left(x_i\right) - \text{Var}\left(x_i - r_i\right)}{\text{Var}\left(x_i\right)}$.

[10] 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 Barn-

rics from the underlying economic behavior of consumers and plans is that we can see just how predictability (by the social planner, by consumers, or by insurers) matters, and then incorporate this into the efficiency metrics.

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.[11] 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 by Einav and Finkelstein and their colleagues.[12] 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 just 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 et al., 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.[13] The fundamental adverse selection problem in plan choice is from inefficient plan pricing.

2.2.1. Demand-based models

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 considered

below).[14] 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 P 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 consumers newly led to buy H when the premium falls to P, *i.e.*, those whose willingness to pay is exactly P. Fig. 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 Fig. 1.[15] 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 Fig. 1 will not fully describe the welfare loss.[16] 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.[17] This is especially important when comparing the efficiency properties of payment systems with different premium structures.

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 spending. Prices for health insurance options varied with geography, creating enough price variation for the authors to estimate how demand and

eveld 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."

[11] 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.

[12] See Einav et al. (2010a,b), Einav and Finkelstein (2011), and Einav et al. (2013).

[13] 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."

[14] Utility and demand models are clearly connected in theory. Einav et al. (2010b) develop a utility basis of demand models, and include a discussion of both types and the associated welfare methods.

[15] 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 Fig. 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.

[16] 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 et al. (2010b), page 326.

[17] As Einav et al. (2010b, 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. . ."
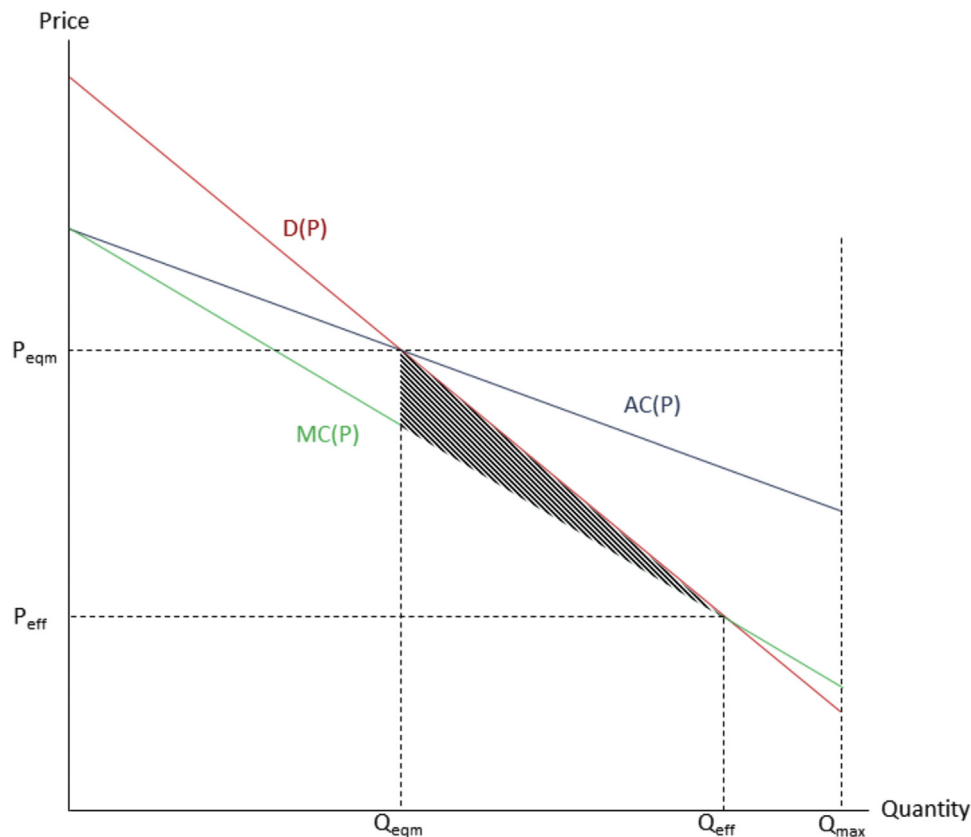
Fig. 1. Efficiency Costs of Adverse Selection in the Einav-Finkelstein Model.

Source: Einav et al. (2010a)

cost varied by price. Both demand and cost curves were downward-sloping in price, in accord with the depiction in Fig. 1 above.[18] 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 et al. (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.[19] 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.[20] More recently, Panhans (2016) used the EF framework to study adverse selection in the Colorado Marketplace, finding that changes to the payment system, specifically the implementation of age-specific subsidies, would result in significant welfare improvements.

EF-type models have also recently been applied to Medicare Advantage (MA). Cabral et al. (2014) use a modified version of the EF framework to estimate the extent of selection into MA, finding little evidence of selection on the margin. Curto et al. (2014) also estimate important structural elements of demand and cost using changes in MA premiums over time, again finding little evidence of selection into MA at the margin (as premiums move up and down)

---

[18] See Einav et al. (2010a) Figure V page 914. The cost curves in the figure are not exactly analogous to the curves estimated by Einav et al. (2010a). In the figure, individuals choose between insurance and no insurance, 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.

[19] 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.

[20] 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 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.

but on average, costs were lower in MA, even after risk adjustment, by 2–3%. More recently, Glazer and McGuire (2017) use the EF conceptual framework to derive the implications for setting the level of subsidy to MA plans.

Finally, a recent set of papers generalize the EF conceptual framework to allow for additional market frictions. First, Mahoney and Weyl (2016) allow for imperfect competition and show that market power has important implications for the consequences of adverse selection. Second, Spinnewijn (2017) and Handel et al. (2015b) generalize the EF framework to allow for behavioral frictions in plan choice, based on previous work by Handel (2013) and Handel and Kolstad (2015) showing that for some consumers behavioral frictions drive a wedge between demand and willingness-to-pay. They use their generalized EF framework to derive implications of these behavioral frictions for the extent of adverse selection and the welfare consequences of policies such as mandates, subsidies, and risk adjustment. Third, Cabral and Cullen (2016) expand the EF framework to allow for estimation of consumer valuation of public insurance programs (such as Social Security Disability Insurance, SSDI) using complementary private insurance. They make the observation that for many public insurance programs, many individuals purchase voluntary supplementary coverage. They show that the EF framework can be used to extrapolate from demand for private coverage to value public coverage, as well as to estimate selection into private coverage.

2.2.2. Utility-based 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 et al. (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).

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 et al. (2015a) use it to study the trade-off between adverse selection and reclassification risk in the Marketplaces. Layton (2017) 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. In another paper, Handel et al. (2015b) use a structural model of insurance demand that accounts for information frictions using consumer responses to survey questions assessing understanding of the plans available to them to show how these information frictions affect the extent of adverse selection and the consequences of risk adjustment.

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. The value of this *ex post* evaluation literature is that it can often estimate key parameters of interest, and evaluate real-world policy changes and consumer behavior, which can provide guidance to policymakers about which types of problems are most important and how well specific policies ameliorate those problems.[21] The major limitation of this *ex post* evaluation literature is precisely that it is *ex post*, implying that the conclusions about the payment system are drawn after the payment system has been designed at which point design changes are much more difficult.

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 Stiglitz (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. Azevedo and Gottlieb (2017) and Veiga and Weyl (2016) generalize the RS model to show that under more general conditions adverse selection always results in some consumers getting less-than-efficient levels of coverage in equilibrium. 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 to attract the good risks (*e.g.* undersupply care that is valued more by sicker people). In Marketplace plans and other settings, nominal coverage is typically regulated, mitigating service-level selection on dimensions like covered services.

---

[21] 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.

However, plans can work around these 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 et al. (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.[22] Building on ideas in Frank et al. (2000), Ellis and McGuire (2007), in an application to Medicare Advantage, show that when plans are designed to maximize profits, services that are predictive and predictable are more attractive for health plans to ration 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 (un)profitability can be altered, mitigating or in some cases eliminating incentives to distort services (Glazer and McGuire, 2002). McGuire et al. (2014) incorporate consideration of demand elasticity into the predictable-predictive framework to study incentives for service-level selection in the Marketplaces with a payment system incorporating risk adjustment. They find significant incentives for insurers to discriminate against individuals with chronic diseases. In a recent paper, Ellis et al. (2017) estimate demand response at a disaggregated level and show that accounting for differences in elasticities of demand across services may be important for accurately assessing service-level selection incentives. 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 et al. (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 non-HMOs. 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 (2017a, 2017b) 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 (2017a) 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 (2017b) 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.[23] Lavetti and Simon (2016) also study selection and drug formulary design, comparing Part D plan formularies to Medicare Advantage-Part D (MA-PD) plan formularies, recognizing that these plans face different selection incentives due to the fact that MA-PD plans cover non-drug medical costs while Part D plans do not. They find that MA-PD plan formularies are consistent with service-level selection distortions and that most of the distortion is driven by enhanced coverage for drugs taken by profitable enrollees rather than worse coverage for drugs taken by unprofitable enrollees.

Additional recent work has focused on identifying evidence of service-level selection among Marketplace plan contracts. Geruso et al. (2016) use data on Marketplace plan and self-insured employer plan formularies to determine whether differences between Marketplace formularies (where selection incentives are strong) and employer formularies (where there are no selection incentives) correspond to the strength and the direction of the selection incentive associated with a particular drug class. They find robust evidence that Marketplace plans severely limit coverage and access for drug classes that are used by the most unprofitable enrollees. Shepard (2015) studies the interaction between adverse selection and hospital network design in the Massachusetts subsidized health insurance exchange, CommonwealthCare, finding that consumers who value the inclusion of a local "star" hospital system most highly also have relatively 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.[24]

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-based 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 each drive a wedge between a consumer's first-best price and the price she is charged in equilibrium. Each source of price distortion leads to Harberger-type measures of welfare loss with the loss proportional to the square of the gap between

---

[22] Geruso and Layton (2017) provide a recent review of this literature.

[23] See also Kuziemko et al. (2014) for a study of Medicaid managed care plans attempting to attract lower cost births based on the race-ethnicity of the mother.

[24] 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 adding the hospital.

the efficient and the equilibrium prices. Both sources of inefficiency emerge within a general model of plan choice.

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 plan, which we will refer to as "gold," is proportional to the cost of enrolling her in the basic ("silver") plan. Although gold and silver are terms referring to Marketplaces, the idea here also apply to consumers choosing between Medicare Advantage and traditional Medicare, and other settings. 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, a concept elaborated on by Geruso (2017). For now, we define type abstractly, but when applying our metrics to data, types need to be 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 $prem_t = prem_t^g - prem_t^s$. In many health insurance markets, such as Medicare Advantage or in those in Israel or the Netherlands, plans set one premium that applies to all enrollees.

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 et al. (1998) and Bundorf et al. (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.

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:

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

In Appendix A of Layton et al. (2015) we show that $\Delta W_t(prem_t)$ can be approximated by the following expression[25]:

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

The approximation indicates 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 $prem_t^* = \gamma x_t$. As shown in Layton et al. (2015), summing over types and assuming that the elasticity of demand for the plan as a function of the incremental premium is the same for all types $(\varepsilon = \frac{n_t'}{N_t})$,[26,27] the welfare loss due to distorted prices in the full population is approximated by:

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

Expression (2) will represent the exact welfare loss in settings where demand is linear (*i.e.*, $n_t''(\gamma x_t) = 0$).[28] 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.[29]

3.1. A health plan payment system: transfers and premiums

When considering efficiency of consumer sorting between plans with fixed characteristics, Expression (2) shows that a payment system affects welfare by moving 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[30]: $trans_{pi} = r_{pi} - \bar{x}$. This structure follows from the typical risk adjustment system that pays insurers the difference between the predicted cost of an enrollee, r_{pi}, 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 $r_{pi}^j = r_{pi}^{j'}$.[31] For example, in a system in which transfers are determined only by risk adjustment, the transfer might be equal

---

[25] The approximation is from a second-order Taylor series expansion of $\Delta W_t(prem_t)$. The second-order approximation is equivalent to assuming that demand $(n_t(prem_t))$ is approximately linear around γx_t.

[26] 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.

[27] 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.

[28] 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.

[29] 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 significant 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, (2) should be a good approximation. 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.

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

[31] 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).

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 $x_{pi} = x_i - trans_{pi} = x_i - (r_{pi} - \bar{x})$ The plan must cover these net costs with premiums.[32]

Plans can set premiums separately for each of a set, $\Gamma_p = \{1, \ldots, a, \ldots, A_p\}$, 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 (1) for welfare loss contains the demand-response parameter $\varepsilon \left(= \frac{n'_t}{N_t} \right)$, 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 system that results in the incremental premium being equal to the average incremental cost in the population, $prem_i = \gamma\bar{x}$ for all i[33]:

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

$$= 1 - \frac{\sum_{i=1}^{N}(prem_{pi} - \gamma x_i)^2}{\sum_{i=1}^{N}(\gamma\bar{x} - \gamma x_i)^2} \qquad (2)$$

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, $prem_{pi}$.

Eq. (2) can be decomposed (see Appendix A in Layton et al. (2015)) into two terms corresponding to the two sources of the price distortions that cause the welfare losses due to inefficient plan choices – the distortion due to limitations on premium groups and the distortion due to adverse selection into the more generous plan. This decomposition aids both in the operationalization of the expression for welfare loss and its interpretation.

$$1 - \frac{\sum_{i=1}^{N}(prem_{pi} - \gamma x_i)^2}{\sum_{i=1}^{N}(\gamma\bar{x} - \gamma x_i)^2} = 1 - [\underbrace{\frac{\sum_{a=1}^{A_p}\sum_{i \in a}(x_i - \bar{x}_a)^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}}_{\tilde{\delta}}$$

$$+ \underbrace{\frac{\sum_{a=1}^{A_p}\sum_{i \in a}(prem_{pi} - \gamma\bar{x}_a)^2}{\gamma^2\sum_{i=1}^{N}(x_i - \bar{x})^2}}_{\tilde{\phi}}] \qquad (3)$$

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 (3), $\tilde{\phi}$, captures the welfare losses due to any deviation of equilibrium prices $prem_{pi}$ 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 Eq. (3), $\tilde{\delta}$, 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.[34] For a single premium, the 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 more premium groups allow incremental premiums to "fit" the distribution of incremental costs better, the measure increases, indicating higher efficiency. At minimum, if premiums fully capture expected incremental cost variation for each individual, premium fit will rise to 1.

3.4. Measuring inefficiency due to adverse selection

We now address $\tilde{\phi}$, the adverse selection component of (3), which characterizes the inefficiency caused by price distortions beyond the distortion due to limited premium categories. This is the distortion studied in EFC (2010) and related papers. We follow EFC (2010) and Handel et al. (2015a) 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.[35] Denoting as $\bar{x}^g_{pa}$ and $\bar{x}^s_{pa}$ the average *silver plan costs* net of transfers of the individuals joining the gold and silver

---

[32] 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 assumption in the empirical section.

[33] 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.

[34] 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 of Layton et al. (2015), 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 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.

[35] 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 of Layton et al. (2015).

plans, respectively, this implies that the gold and silver premiums for individuals in premium category z will be equal to[36]:

$$\text{prem}_{pa}^g = (1 + \gamma)\bar{x}_{pa}^g \quad \text{prem}_{pa}^s = \bar{x}_{pa}^s$$

The equilibrium incremental premium for each premium category, prem_a^e, must then be equal to the difference in these average plan net costs[37]:

$$\text{prem}_{pa}^e = \text{prem}_{pa}^g - \text{prem}_{pa}^s = (1 + \gamma)\bar{x}_{pa}^g - \bar{x}_{pa}^s \quad (4)$$

We can now plug our expression for premiums from Eq. (4) into the expression for ϕ from Eq. (3) to get:

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

This measure is effectively a weighted sum of the difference between each premium group's equilibrium price from its second-best price. These differences are small when transfers due to risk adjustment or other payment system features track costs across the premium groups. $\tilde{\phi}$ is a fit-like measure that, with certain regularity assumptions can be simplified to[38]:

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

where ω is a scaling parameter that depends on γ and properties of the variation in costs but is independent of the payment system.

Dividing by ω (since it is the same in all payment system alternatives), we define $1 - \phi = 1 - \frac{\tilde{\phi}}{\omega}$ 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.*, $\sum_{i=1}^{N}[x_i - (r_{pi} + (\bar{x}_a - \bar{r}_{pa}))]^2 = \sum_{i=1}^{N}[x_i - \bar{x}]^2$), 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.*, $\sum_{i=1}^{N}[x_i - (r_{pi} + (\bar{x}_a - \bar{r}_{pa}))]^2 = 0$), 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 choice. The first, $1 - \delta$, measures the inefficiency due to limited premium categories, 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.

We note that premium fit is important independent of payment system fit. Even if the payment system perfectly matches revenues to costs, thereby making payment system fit equal to 1, inefficiencies typically remain. The opposite is not true, however. If premium fit is equal to 1, payment system fit is also equal to 1 and all selection-related inefficiencies are eliminated. The independent importance of premium fit was previously pointed out by Bundorf et al. (2012) and Geruso (2017). Practical importance of this insight for policy evaluation may be limited, however. In many applications, regulation of premium categories is not subject to change. In these circumstances, the sorting efficiency issue is about adverse selection into a more generous plan, an EF-type of problem, and payment systems can be compared using only ϕ and not δ.

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 present a measure of welfare loss due to inefficient allocations of health care spending across people and services developed in our earlier work (Layton et al., 2016b). The measure generalizes earlier approaches in that it incorporates both discrimination based on groups of 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). Welfare loss is driven by the wedge between the efficient allocation to an individual and the allocation the individual would receive in equilibrium under a given health plan payment system. The measure thus applies to inefficiencies related to the services offered by health plans, and not to inefficiencies related to advertising or other plan actions distinct from the distortion of the health insurance contract itself and the benefits and costs of health care under that contract. Throughout, we maintain the assumption that health plans are profit maximizers and compete in a market.

4.1. A Health insurance contract

In this section, we describe the development of our measure of welfare loss due to inefficient allocations of health care spending in Layton et al. (2016b). 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. Services can be defined as broadly or as narrowly as the researcher or policymaker desires, but they should capture all dimensions of spending that an insurer can feasibly distort. A

---

[36] 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.

[37] 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.

[38] A set of sufficient assumptions 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. These assumptions are explicitly laid out in Appendix A of Layton et al. (2015).

health plan can thus be described as an $N \times S$ matrix $\mathbf{X}$ of individual allocations of medical services:

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

Total spending on person i for all services is $x_i = \sum_s x_{is}$. We define an individual's valuation of service s as $v_{is}(x_{is})$, with $v'_{is}(x_{is}) > 0$ and $v''_{is}(x_{is}) < 0$. We also denote the first-best level of x_{is} as x^*_{is} so that $v'_{is}(x^*_{is}) = 1$. Finally, we denote the level of x_{is} the insurer offers in the equilibrium contract as x^e_{is}, so that $x^e_i = \sum_s x^e_{is}$ and

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

While in Section 3, welfare losses occur due to price distortions, here welfare losses are due to quantity distortions where adverse selection causes insurers to offer equilibrium contracts that differ from the first-best contract. More formally, net welfare for individual i under equilibrium contract $\mathbf{X^e}$ is $W_i(\mathbf{X^e}) = \sum_s v_{is}(x^e_{is}) - x^e_i$ and welfare loss relative to the first-best is $\delta W_i(\mathbf{X^e}) = [\sum_s v_{is}(x^e_{is}) - x^e_i] - [\sum_s v_{is}(x^*_{is}) - x^*_i]$. Similar to Section 3, a (second-order) Taylor-series expansion of $\Delta W_i(\mathbf{X^e})$ around x^*_{is} yields

$$\delta W_i(\mathbf{X^e}) \approx \frac{1}{2} \sum_s v''_{is}(x^*_{is})(x^e_{is} - x^*_{is})^2$$

which can be summed across the entire population to produce

$$\delta W(\mathbf{X^e}) \approx \frac{1}{2} \sum_i \sum_s v''_{is}(x^*_{is})(x^e_{is} - x^*_{is})^2 \quad (5)$$

Thus, welfare losses under contract $\mathbf{X^e}$ are proportional to the weighted sum of squared differences between the equilibrium and the first-best allocations where the weight is the second-derivative of the individual's service-specific valuation function.

While first-best allocations, x^*_{is}, are fixed, equilibrium allocations, x^e_{is}, will depend on the payment system. We now lay out the necessary assumptions to draw an explicit connection between the revenues generated for person i by a given payment system, and the equilibrium allocations. This explicit connection will allow us to replace x^e_{is} in Eq. (5) with an expression related to the revenues generated by the payment system, leading to a welfare-based measure of payment system performance.

A payment system generates a payment r_i for person i. We assume that the payment system is structured such that r_i is equal to the linear combination of a set of individual characteristics, z_{ik}, and corresponding payment system weights, β_k: $r_i = \sum_k \beta_k z_{ik}$. We note that this definition of revenues is general and allows for the possibility of various forms of risk-sharing such as reinsurance, as

shown by Layton and McGuire (2017), as well as premium groups, as shown by McGuire et al. (2014).[39]

We assume that insurers can modify the overall levels of spending on a given health care service (*i.e.* total spending on diabetes-related care), but that they cannot choose how that total level of spending is allocated across individuals. Instead, the allocation across individuals is fixed. To formalize this notion, we introduce a parameter σ_{is}, with $\sum_s \sigma_{is} = 1$, that defines the share of total spending on service s allocated to individual i such that $x_{is} = \sigma_{is} x_s$ where x_s is the total spending on service s across all consumers, $x_s = \sum_i x_{is}$. The service spending shares, σ_{is}, are assumed to be fixed across payment systems.

We now consider the insurer's problem of choosing total spending on each service, x_s. We assume that insurers choose x_s to maximize expected profits. If we define $Pr_i^j(\mathbf{X^j})$ as the probability that person i enrolls in plan j given a contract $\mathbf{X^j}$, then plan profits can be written as

$$\pi = \sum_i Pr_i^j(\mathbf{X})(r_i - x_i)$$

We place some further assumptions on $Pr_i^j(\mathbf{X^j})$. Specifically, we assume that $Pr_i^j(\mathbf{X^j}) = Pr_i(v_i(\hat{\mathbf{X}}^j)) = Pr_i(\sum_s v_{is}(\hat{x}^j_{is}))$. There are a number of key assumptions here. The first is that the probability that person i enrolls in plan j depends on her valuation of the contract. The second is that i's valuation of the contract is the sum of her valuation of each service-level allocation. The third is that i cares not about the *actual* allocations, x_{is}, but about her *expected* allocations, which we denote $\hat{x}_{is}$. Given these assumptions, we find that the first order condition of plan profits with respect to x_s is

$$\sum_i Pr_i'(v_i(\hat{\mathbf{X}}))v'_{is}\hat{\sigma}_{is}(r_i - x_i) - \sum_i Pr_i(v_i(\hat{\mathbf{X}}))\sigma_{is} = 0$$

Where $\hat{\sigma}_{is}$ represents person i's expected share of total plan spending on service s. Letting $\alpha_{is} = Pr_i'(v_i(\hat{\mathbf{X}}))v'_{is}\hat{\sigma}_{is}$, it is straightforward to show that this implies that under plan profit maximization, the following $S-1$ equations must hold:

$$\sum_i \alpha_{is}(r_i - x_i) = \sum_i \alpha_{is'}(r_i - x_i)$$

Now, by invoking the same perfect competition assumption used in Section 3 that implies that total plan spending be equal to total plan revenues (*i.e.* the zero profit condition), $\sum_i r_i = \sum_i \sum_k \beta_k z_{ik} = \sum_i x_i$, we have S equations with S unknowns (the x_ss), and we can thus solve for the equilibrium levels of overall service-level spending, x^e_s, which then also give us the equilibrium levels of individual-level service-level spending, $x^e_{is} = \sigma_{is} x^e_s$. As we show in Layton et al. (2016b), these equilibrium overall service-level spending levels can be written as the linear combination of the payment system weights, β_k, and a straightforward transformation and aggregation of the variables describing the individual

---

characteristics upon which payments are based, which we denote $\tilde{z}_{sk}$:

$$x_{is}^e = \sigma_{is} x_s^e = \sigma_{is} \sum_k \beta_k \tilde{z}_{sk}$$

This can be plugged into Eq. (5) above to produce an expression for the welfare loss under a payment system where payments are based on individual characteristics, z_{ik}, and their corresponding payment weights, β_k:

$$\delta W(x_i^e) = \frac{1}{2} \sum_i \sum_s v_{is}''(\sigma_{is} \sum_k \beta_k \tilde{z}_{sk} - x_{is}^*)^2 \tag{6}$$

This expression leads to a natural measure, analogous to the measures developed in Section 3, of welfare loss under a given payment system relative to welfare loss under a payment system where revenues are equal to the average cost in the full population for all enrollees (*i.e.*, community-rated premiums across the market with no risk adjustment or reinsurance). To get to the final measure, we first make the assumption that $v_{is}'' = v''$ for all i and s. We then define $\tilde{z}_{sk}^{nora}$ as the transformed version of z_{sk} for the payment system where revenues for all individuals are equal to the average cost in the full population (see Layton et al. (2016b) for details). This allows us to define the relative welfare loss measure as

$$\psi = 1 - \frac{\frac{1}{2}v'' \sum_i \sum_s (\sigma_{is} \sum_k \beta_k \hat{z}_{sk} - x_{is}^*)^2}{\frac{1}{2}v'' \sum_i \sum_s (\sigma_{is} \hat{z}_s^{nora} - x_{is}^*)^2}$$

$$= 1 - \frac{\sum_i \sum_s (\sigma_{is} \sum_k \beta_k \hat{z}_{sk} - x_{is}^*)^2}{\sum_i \sum_s (\sigma_{is} \hat{z}_s^{nora} - x_{is}^*)^2}$$

This measure, like the measures in Section 3, mimics the form of the classic R-squared metric of statistical fit. When the measure is equal to 1, the welfare loss due to inefficient spending allocations related to adverse selection is completely eliminated. When the measure is equal to zero, the payment system being evaluated performs equally to a completely neutral payment system under which the revenues for each individual are equal to average cost. The intuition behind the measure is that the payment system is assessed by how close the payment system brings equilibrium individual-by-service spending ($\sigma_{is} \sum_k \beta_k \hat{z}_{sk}$ in our model) to first-best spending x_{is}^*. This differs from the traditional R-squared metric which assesses payment system performance based on how well individual-level revenues "fit" the variation in individual-level costs.[40]

4.2. Summary of Selection Measures

Table 1 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 payment system features (*i.e.* premium categories and other features such as risk adjustment and/or reinsurance) 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 expression for the measure derived in Section 4, the measure of efficiency loss due to inefficient distortions of plan benefits caused by adverse selection. To operationalize this measure the researcher or policymaker must choose a service categorization. Some additional assumptions are required in order to make this measure operational using *ex ante* data. We discuss these assumptions in our illustrative application below.

The measures vary between 0 and 1 with 0 implying the payment system fails to improve efficiency in relation to a base payment system with equal payment for all enrollees and 1 implying complete elimination of inefficiency. Notably, the measures have affinities to the statistical measures traditionally used in the literature to assess 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 (premium regulations make first-best prices and sorting impossible) and the payment system fit is a metric relevant to EF-type selection problems (adverse selection causes equilibrium prices to diverge even from the second-best prices that sort consumers as efficiently as possible given premium regulations). The third measure is similar to an aggregation of an extended version of the service-level selection index developed by Ellis and McGuire (2007) across services. It assesses payment system performance based on the squared difference between first-best spending levels and the equilibrium spending levels derived under our model. We will compute the second and third measures of ex ante payment system performance for policy alternatives for Marketplaces in 2017.

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 both years,[41] 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. After applying the inclusion and exclusion criteria, we have two years of claims for 7,072,964 individuals. We then further restrict the data to a subset of around 2 million individuals that look similar to individuals who are eligible for Marketplace coverage based on a set of observable characteristics. We refer to this subset as the MarketScan Marketplace (MM) sample. More details about our methods for implementing this restriction can be found in Layton et al. (2015).

---

[40] While our measure clearly captures the economic goals of interest, the R-squared instead focuses on a statistical goal that only aligns with the economics goals in a very special case outlined in Layton et al. (2016b).

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

Table 1
Summary of Selection Measures.

	Measure	Description
Premium fit	$\delta = 1 - \dfrac{\sum_i (x_i - \bar{x}_a)^2}{\sum_i (x_i - \bar{x})^2}$	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	$\phi = 1 - \left[\dfrac{\sum_{i=1}^{N} (x_i - (r_{p,i} + (\bar{x}_a - \bar{r}_{p,a})))^2}{\sum_{i=1}^{N} (x_i - \bar{x})^2} \right]$	R-squared from regression producing predicted values equal to $(r_{p,i} + (\bar{x}_a - \bar{r}_{p,a}))$. Based on expected spending from the insurer's point of view. Captures Einav-Finkelstein type inefficiency.
Welfare-based metric of incentives to distort benefits	$\psi = 1 - \dfrac{\sum_i \sum_s (\sigma_{is} \sum_k \beta_k \hat{z}_{sk} - x_{is}^*)^2}{\sum_i \sum_s (\sigma_{is} \hat{z}_s^{nora} - x_{is}^*)^2}$	R-squared-like measure incorporating the difference between "equilibrium" spending and "efficient" spending at the individual-by-service level. 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. 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 2
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	$5059	$20,134	$5838	$18,779
Inpatient Spending	$1357	$13,657	$1221	$10,910
Outpatient Spending	$2779	$10,204	$3344	$11,150
Drug Spending	$922	$4189	$1273	$5251
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.

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 2. Most of the variables are self-explanatory. 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 studied

We apply our measures to two payment system alternatives. The first 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 (CCIIO, 2016). While states and the federal government may modify this payment system in future years, it still represents a relevant baseline for comparing alternative payment systems.

The second, alternative system, 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 and beyond, as diagnoses from prior year claims are now available.[42] As with the ACA 2017 payment system,

---

[42] In Layton et al. (2015) we assume that the regulator does not have sufficient information to use prospective risk adjustment for a randomly selected 50% of the population. Here, for simplicity, we do not do this.

we maintain the HHS transfer formula to avoid over-adjusting for age. This alternative also restores reinsurance, paying 80% of costs over an attachment point of $60,000 in plan costs.[43]

5.3. Estimation of selection measures

5.3.1. Measures related to sorting between plans

The selection measures from Section 3 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 the medical spending observed in the data. In Layton et al. (2015) we relax this assumption and show that most findings are not sensitive to alternative specifications of expected costs. r_{pi} and rev_{pi} (transfers and revenues) must be simulated under each payment system, taking into account premium differentiation, risk adjustment, reinsurance, and the risk adjustment transfer formula. We simulate transfers and revenues using the HHS-HCC risk adjustment system used in the Marketplaces (including the federal transfer formula used to determine risk adjustment transfers), a budget neutral version of the reinsurance policy initially proposed for the Marketplaces, and the federal age curve used to determine age-specific premiums. Details on the simulation of transfers and revenues for the two payment systems we assess in this paper are provided in Layton et al. (2015).

5.3.2. Measures related to plan design

In order to estimate the measure of welfare loss due to service-level distortions from Section 4 we also have to provide additional definitions and assumptions. First, we have to choose a partition of spending into services. Typically, this partition should be at the level at which insurers can feasibly modify spending. Because this is just an illustration of the use of these measures in the evaluation of payment system performance, we adopt an extremely simple partition of health care services: inpatient, outpatient, and prescription drugs.

In order to construct the measure, we also require a measure of "first-best" service-level spending (x_{is}^* above), the parameters dictating the portion of total spending on service s that is allocated to individual i (σ_{is}), the actual payment weights for each payment system, β_{pk}, and the parameter $\alpha_{is} = Pr_i'(v_i(\hat{X}))v_{is}'\hat{\sigma}_{is}$ from Section 4 above. Additionally, because reinsurance is a component of one of the payment systems we simulate, we need a method for incorporating reinsurance into the measure.

We start by assuming that the first-best levels of spending are equal to the levels of spending observed in the data, *i.e.* if individual i consumed $500 of service s in the MM dataset, $x_{is}^* = 500$. This assumption may seem odd but, as we argue in Layton et al. (2016b), it is an assumption that is implicit in most of the papers in the literature evaluating health plan payment systems. We also refer the interested reader to our lengthy discussion of this issue in Layton et al. (2016b).

Next, we make an assumption about σ_{is}. Again, we follow our previous work by assuming that σ_{is} (and $\hat{\sigma}_{is}$) is fixed across all relevant service-level allocations. This assumption implies that the σ_{is} is equal to the share of spending on service s consumed by individual i observed in the data. Again, we refer the reader to Layton et al. (2016b) for a more comprehensive discussion of this assumption.

Next we turn to the payment weights and the α parameters. The payment weights for each payment system, β_{pk}, are estimated *via* linear regression as discussed in Section 5.3.1. For the α parameters we make an assumption that effectively implies that for each service consumers respond about equally in plan choice

to changes in allocations of spending. Formally, we assume that $Pr_i'(v_i(\hat{X}))v_{is}' = \alpha_{is} = \alpha$.[44]

The derivation of our measure of welfare losses due to service-level distortions in Section 4 assumes that revenues are equal to a linear combination of risk adjusters and payment weights. In our simulations, we incorporate reinsurance which, at first glance, does not follow the linear combination formulation. However, as shown in Layton and McGuire (2017) reinsurance payments can be incorporated into a typical linear risk adjustment formula.

6. Results

This section presents estimates of our selection measures and compares them to the commonly applied R-squared and predictive ratio measures. Robustness and sensitivity analyses for the measures can be found in Layton et al. (2015) and Layton et al. (2016b).

6.1. Basic results

Fig. 2 displays our first two measures of efficiency for the ACA 2017 payment system and the alternative payment system incorporating prospective risk adjustment plus reinsurance. 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. Recall that this measure captures the extent to which consumers' equilibrium premiums can possibly equal their first-best premiums even in the absence of adverse selection. which captures the inefficiency due to premium regulation. The measure is equal to 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 of Layton et al. (2015)). The low level of the dark bars indicate that the government-issued age curve doesn't do much in the way of matching premium categories to costs. The ACA 2017 payment system and the alternative payment system have identical premium fit because both 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.[45] Because the reinsurance and risk adjustment payments modeled here are all budget neutral, the entire age structure of premiums is the same under these two payment systems. For both the ACA 2017 system and the alternative system, the premium fit is precisely estimated to be 0.013.

The light bars in the figure represent payment system fit, where there is 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).[46] Payment system fit for the alternative payment system, prospective risk adjustment plus rein-

---

[44] For additional discussion of this assumption see footnote 23 in Layton et al. (2016b).

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

[46] 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
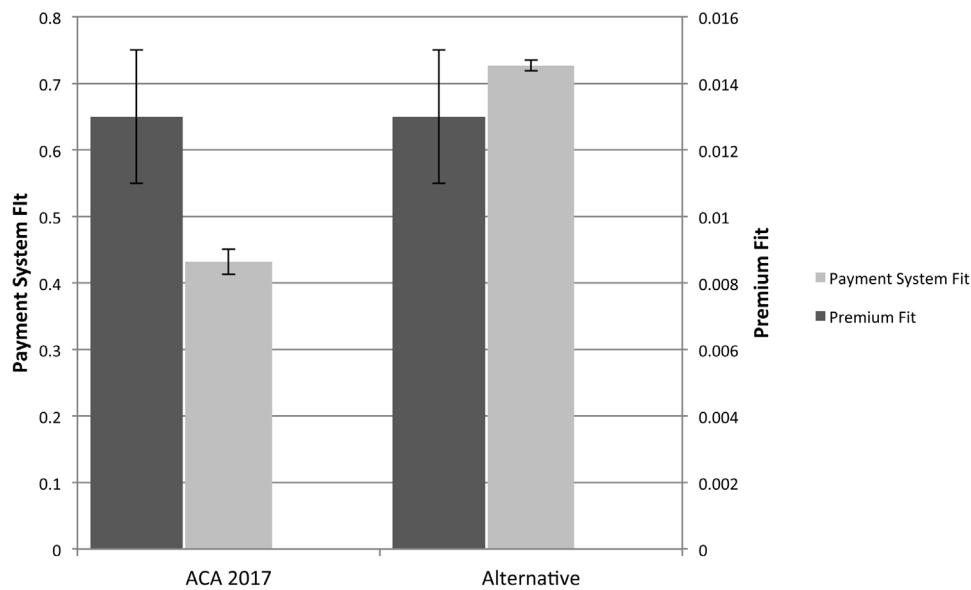
Fig. 2. Premium Fit and Payment System Fit for Each Simulated Payment System.
Notes: The figure displays premium fit (right y-axis) and payment system fit (left y-axis), the measures of welfare loss due to price distortions caused by premium regulations and adverse selection derived in Section 3.1 implies 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 + federal age curve + reinsurance.

Table 3
Estimated Measures for Simulated Payment Systems.

Measure	ACA 2017	Alternative
Payment System Fit: Φ	0.423	0.727
	(0.414, 0.433)	(0.711, 0.743)
Incentives to Distort Benefits:	0.944	0.941

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. The measure of incentives to distort benefits captures the welfare loss from benefit distortions. Services are defined as inpatient, outpatient, and prescription drugs. ACA 2017 is concurrent risk adjustment + federal age curve; Alternative is Prospective risk adjustment (with 50% "new enrollees" assigned age/gender risk scores) + federal age curve + reinsurance.

surance is much larger, 0.727, implying greater efficiency. The estimates of payment system fit are also found in Table 3. In Layton et al. (2015) we show that the result that a payment system incorporating prospective risk adjustment plus reinsurance performs better on payment system fit than a system featuring concurrent risk adjustment alone is largely robust to various assumptions regarding the portion of spending that is "predictable" to the insurer.[47]

Fig. 3 presents the measure developed in Section 4 and in Layton et al. (2016b) that capture welfare losses due to benefit distortions caused by adverse selection. These measures also range between zero (no inefficiency reduction) and one (eliminating inefficiency). The results here differ from the payment system fit results: here, the ACA 2017 payment system outperforms the alternative system consisting of prospective risk adjustment plus reinsurance, though only

slightly. The ACA 2017 system produces a measure of 0.944, while the alternative system produces a measure of 0.941. This implies that with respect to benefit distortions, concurrent risk adjustment alone is at least as effective as prospective risk adjustment plus reinsurance, differing from the conclusions about payment system performance drawn from payment system fit above.[48] The estimates of this measure for the two payment system alternatives are also presented in Table 3.

6.2. Comparison to conventional metrics

To illustrate the discrepancy between outcomes in terms of statistical fit of a risk adjustment model and outcomes in terms of economic efficiency of an entire payment system we will now compare our new metrics with the conventional R-squared for the risk adjustment model and predictive ratios for selected disease categories. The R-squared measures are presented in the first row of Table 4. The ordinary R-squared for concurrent risk adjustment is slightly larger than our estimate of payment system fit for the ACA 2017 payment system. In this setting 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 of Layton et al. (2015), 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 $\frac{1}{N}\sum_{i=1}^{N}(\frac{\alpha_i}{\bar{\alpha}})^2$. 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.

---

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.

[47] Additionally, we refer the interested reader to Layton et al. (2015) for a detailed discussion about where to use *ex ante* "expected" spending rather than *ex post* realized spending.

[48] Note that this differs from the results in Layton et al. (2015) where a different measure of insurer incentives to distort benefits is used.
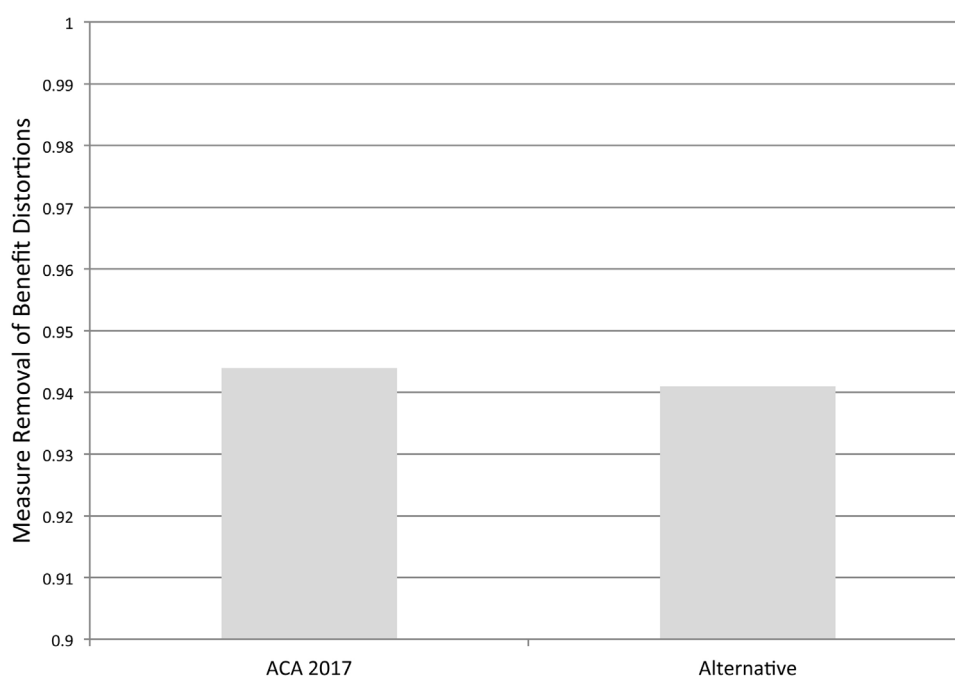
Fig. 3. Measure of Removal of Benefit Distortions.
Notes: The figure displays the measure of inefficiency due to benefit distortion, derived in Section 4 of the paper and in Layton et al. (2016b). The measure describes inefficiency of the given payment system relative to a base payment system with no transfers and a single premium. 1 implies complete elimination of the inefficiency and 0 implies no better than the base payment system. The measure requires the researcher to divide spending into mutually exclusive services. For this illustration of the measure, we define services as inpatient, outpatient, and prescription drug. The simulated payment systems are as follows: ACA 2017, concurrent risk adjustment + federal age curve; Alternative, Prospective risk adjustment + federal age curve + reinsurance.

Table 4
Conventional Measures for Simulated Payment Systems.

Measure	ACA 2017	Alternative
R-squared	0.432	0.079
Payment System Predictive Ratios:		
Cancer	0.920	0.797
Diabetes	0.956	0.820
Heart Disease	0.825	0.716
Mental Health	0.807	0.704
No Chronic Condition	1.270	1.538

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 two simulated payment systems are shown: ACA 2017 is concurrent risk adjustment + federal age curve; Alternative is Prospective risk adjustment + federal age curve + reinsurance.

For the alternative payment system featuring prospective risk adjustment and reinsurance, the difference between the R-squared for the risk adjustment model 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. For those interested in overall payment system fit the conventional R-squared measure can be seriously misleading when a payment system incorporates features other than regression-based risk adjustment.

Fig. 4 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 mea-

sure of welfare losses due to benefit distortions. 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 Fig. 4 for the ACA 2017 payment system. For the alternative system 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.

Consistent with the findings for our measure of welfare losses due to distorted benefits, all conventional predictive ratios are better under the ACA 2017 payment system than under prospective risk adjustment plus reinsurance. 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 4.

7. Discussion

This paper develops metrics for evaluating payment system performance with respect to selection-related inefficiencies that we argue are valid, complete and practical to use for the *ex ante* evaluation 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
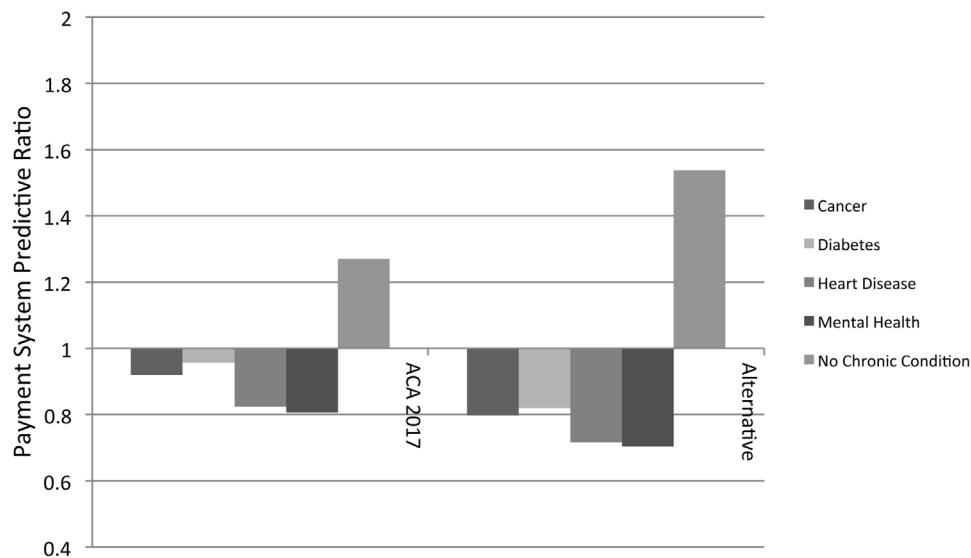
Fig. 4. Payment System Predictive Ratios.
Notes: The figure presents payment system predictive ratios for various disease groups under each of the simulated payment systems. Payment system predictive ratios are defined as average revenues divided by average costs for the specified group under the given 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. 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, Prospective risk adjustment + federal age curve + reinsurance.

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 2018, 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 actions to distort benefits to attract profitable enrollees. This measure requires that the researcher or policymaker specify the "services" that insurers can manipulate *via* benefit design. The measure is then based on the difference between the efficient level of spending and the equilibrium level of spending for a particular service for a particular individual. This difference is determined by a model of benefit design by a profit maximizing insurer in a perfectly competitive health insurance market, and it is related to the "predictability and predictiveness" of the service as originally pointed out by Ellis and McGuire (2007).

Before summarizing the findings of the paper, we provide a brief "instruction manual" for how to use these measures. First, we note that premium fit will always be identical for payment systems that have the same premium groups. Thus, when comparing such payment systems, this measure can be ignored. Such comparisons are common in the economics and policy literature. Payment system fit, on the other hand, is always a relevant metric to use to compare payment systems. The single requirement for computing payment system fit is that the researcher generate simulated health plan revenues at the individual level including all features of the payment system (premiums, risk adjustment, risk sharing). Our measure of welfare losses due to benefit distortions is also always relevant. This measure is slightly more onerous, however, in that it requires

that the researcher both generates simulated revenues *and* specifies a partition of healthcare spending into a mutually exclusive set of services that insurers can manipulate *via* benefit design.

We demonstrated the usefulness of our measures by comparing an 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 plus reinsurance performs notably better than the proposed concurrent model on payment system fit, this result does not necessarily carry over to our measure of benefit distortions. This result regarding the value of reinsurance for payment system fit is consistent with other recent research bearing on reinsurance in the Marketplace context (Zhu et al., 2013; Geruso and McGuire 2016), and it is robust to assumptions regarding the portion of the variance of medical spending that is "predictable" by consumers (see Layton et al. (2015) and Layton et al. (2016b)). The result regarding equal performance with respect to benefit distortions is new and should be explored in greater depth.

Our approach is subject to a number of limitations. The most important is that our measures rely on a number of strong assumptions about insurer and consumer behavior. It is important to recognize, however, that the measures typically used to evaluate payment system performance also implicitly rely on strong assumptions, we have just made them more explicit (see Layton et al. (2016b) for a more detailed discussion of this point). Moreover, 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, under the existing Marketplace payment system, 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 other potentially inefficient insurer incentives embedded in the payment system. One example of an inefficient insurer incentive is the "power" of the payment system, a concept originally introduced by Geruso and McGuire (2016) that refers to insurer incentives for cost control. They consider both power and payment system fit simultaneously, finding that some systems "dominate" others in that they have higher power and higher payment system fit. Specifically, their results suggest 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.

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 et al., 2015a). 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. Q. J. Econ. 84 (3), 488–500.

Azevedo, Eduardo, Gottlieb, Daniel, 2017. Perfect competition in markets with adverse selection. Econometrica 85 (1), 67–105.

Beck, K., Trottman, M., Zweifel, P., 2010. Risk adjustment in health insurance and its long-Term consequences. J. Health Econ. 29 (4), 489–498.

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., Heineck, M., Lorenz, N., 2003. Determinants of health care utilization by german sickness fund members-with application to risk adjustment. Health Econ. 12 (5), 367–376.

Breyer, F., Bundorf, K., Pauly, M.V., 2012. Health care spending risk, health insurance, and payment to health plans. In: Pauly, M., McGuire, T., Barros, P. (Eds.), The Handbook of Health Economics. Elsevier, p. 2.

Brown, J., Duggan, M., Kuziemko, I., Woolston, W., 2014. How does risk selection respond to risk adjustment? Evidence from the Medicare Advantage program. Am. Econ. Rev. 104 (10), 3335–3364.

Bundorf, M.K., Levin, J.D., Mahoney, N., 2012. Pricing and welfare in health plan choice. Am. Econ. Rev. 102 (7), 3214–3248.

Center For Consumer Information & Insurance Oversight (CCIIO), 2016. Discussion Paper Prepared for the HHS-Operated Risk Adjustment Methodology Meeting, March 31, 2016. CCIIO, Centers for Medicare & Medicaid Services.

Cabral, M., Cullen, M., 2016. Estimating the Value of Public Insurance Using Complementary Private Insurance. NBER, Working Paper No. w22583.

Cabral, M., Geruso, M., Mahoney, N., 2014. Does privatized health insurance benefit patients or producers? Evidence from Medicare Advantage. NBER, Working Paper 20470.

Cao, Z., McGuire, T., 2003. Service-level selection by HMOs in medicare. J. Health Econ. 22 (6), 915–931.

Carey, C., 2017a. Technological change and risk adjustment: benefit design incentives in medicare part D. Am. Econ. J. Econ. Policy 9 (1), 38–73.

Carey, C., 2017b. Time to Harvest: Evidence on Consumer Choice Frictions from a Payment Revision in Medicare Part D, Working Paper, Retrieved from https://drive.google.com/file/d/0B2TeS7lispKBQ21RLV9PckE2eWs/view.

Cohen, A., Einav, L., 2007. Estimating risk preferences from deductible choice. Am. Econ. Rev. 97 (3), 745–788.

Curto, V., Einav, L., Levin, J., Bhattacharya, J., 2014. Can Health Insurance Competition Work? Evidence from Medicare Advantage. National Bureau of Economic Research, Working Paper 20818, December.

Cutler, D.M., Reber, S.J., 1998. Paying for health insurance: the tradeoff between competition and adverse selection. Q. J. Econ. 113 (2), 433–466.

Eggleston, K., Bir, A., 2009. Measuring selection incentives in managed care: evidence from the Massachusetts state employees insurance program. J. Risk Insur. 76, 159–175.

Einav, L., Finkelstein, A., 2011. Selection in insurance markets: theory and empirics in pictures. J. Econ. Perspect. 25 (1), 115–138.

Einav, L., Finkelstein, A., Cullen, M.R., 2010a. Estimating welfare in insurance markets using variation in prices. Q. J. Econ. 125 (3), 877–921.

Einav, L., Finkelstein, A., Levin, J.1, 2010b. Beyond testing: empirical models of insurance markets. Ann. Rev. Econ. 2, 311–336.

Einav, L., Finkelstein, A., Ryan, S., Schrimpf, P., Cullen, M., 2013. Selection on moral hazard in health insurance. Am. Econ. Rev. 103 (1), 178–219.

Ellis, R.P., McGuire, T.G., 2007. Predictability and predictiveness in health care spending. J. Health Econ. 26 (1), 25–48.

Ellis, R.P., Jiang, S., Kuo, T.-C., 2013. Does service-level spending show evidence of selection across health plan types. Appl. Econ. 45 (13), 1701–1712.

Ellis, R.P., Martins, B., Zhu, W., 2017. Demand Elasticities and Service Selection Incentives Among Competing Private Health Plans. Boston University working paper, Unpublished.

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

Ettner, S., Frank, R., McGuire, T., Hermann, R., 2001. Risk adjustment alternatives in paying for behavioral health care under medicaid. Health Serv. Res. 36 (4), 793–811.

Frank, R.G., Glazer, J., McGuire, T.G., 2000. Measuring adverse selection in managed health care. J. Health Econ. 19 (6), 829–854.

Geruso, M., Layton, T., 2015. Upcoding: Evidence from Medicare on Squishy Risk Adjustment. NBER, Working Paper, pp. 22832.

Geruso, M., Layton, T., 2017. Selection in insurance markets: policy through the lenses of fixed and endogenous contracts. J. Econ. Perspect., forthcoming.

Geruso, M., McGuire, T.G., 2016. Tradeoffs in the design of health plan payment systems: fit, power and balance. J. Health Econ. 47, 1–19.

Geruso, M., Layton, T., Prinz, D., 2016. Screening in Contract Design: Evidence from the ACA Health Insurance Exchanges. NBER, Working Paper 21222.

Geruso, Michael, 2017. Demand heterogeneity in insurance markets: implications for equity and efficiency. Quan. Econ. (forthcoming).

Glazer, J., McGuire, T.G., 2000. Optimal risk adjustment of health insurance premiums: an application to managed care. Am. Econ. Rev. 90 (4), 1055–1071.

Glazer, J., McGuire, T.G., 2002. Setting health plan premiums to ensure efficient quality in health care: minimum variance optimal risk adjustment. J. Public Econ. 84 (2), 153–173.

Glazer, J., McGuire, T.G., 2017. Paying medicare advantage plans: to level or tilt the playing field. J. Health Econ. 56, 281–291.

Hackmann, M.B., Kolstad, J.T., Kowalski, A.E., 2015. Adverse selection and an individual mandate: when theory meets practice. Am. Econ. Rev. 105 (3), 1030–1066.

Handel, B., Kolstad, J.T., 2015. Health insurance for 'Humans': information frictions, plan choice, and consumer welfare. Am. Econ. Rev. 105 (8), 2449–2500.

Handel, B., Hendel, I., Whinston, M.D., 2015a. Equilibria in health exchanges: adverse selection vs reclassification risk. Econometrica 83 (4), 1261–1313.

Handel, B., Kolstad, J., Spinnewijn, J., 2015b. Information Frictions and Adverse Selection: Policy Interventions in Health Insurance Markets. NBER, Working Paper 21759.

Handel, B., 2013. Adverse selection and inertia in health insurance markets: when nudging hurts. Am. Econ. Rev. 103 (7), 2643–2682.

Hsu, J., Fung, V., Huang, J., et al., 2010. Fixing flaws in medicare drug coverage that prompt insurers to avoid low-income patients. Health Aff. (Millwood) 29 (12), 119–141.

Kautter, J., Pope, G.C., Ingber, M., Freeman, S., Patterson, L., Cohen, M., Keenan, P., 2014. The HHS-HCC risk adjustment model for individual and small group markets under the affordable care act. Med. Med. Res. Rev. 4 (3), E1–E11.

Keeler, E., Carter, G., Newhouse, J., 1998. A model of the impact of reimbursement schemes on health plan choice. J. Health Econ. 17 (3), 297–320.

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

Kronick, R., Welch, P.W., 2014. Measuring coding intensity in the medicare advantage program. Med. Med. Res. Rev. 4 (2), E1–E19.

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.

Lavetti, K., Simon, K., 2016. Strategic formulary design in Medicare Part D plans. Natl. Bureau Econ. Res., Working Paper 22338.

Layton, T., McGuire, T.G., 2017. Marketplace plan payment options for dealing with high-cost enrollees. Am. J. Health Econ. 3 (No. 2), 165–191.

Layton, T., Ellis, R., McGuire, T., 2015. Assessing Incentives for Adverse Selection in Health Payment Systems. National Bureau of Economic Research, Working Paper 21531, September.

Layton, T., McGuire, T.G., Sinaiko, A.S., 2016a. Risk corridors and reinsurance in health insurance marketplaces: insurance for insurers, american. J. Health Econ. 2 (1), 66–95.

Layton, T., McGuire, T., Van Kleef, R., 2016b. Deriving Risk Adjustment Payment Weights to Maximize Efficiency of Health Insurance Markets. National Bureau of Economic Research, Working Paper 22642, September.

Layton, T., 2017. Imperfect risk adjustment, risk preferences, and sorting in competitive health insurance markets. J. Health Econ. (forthcoming).

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.

Mahoney, N., Weyl, E., 2016. Imperfect Competition in Selection Markets (June 8, s2016). Rev. Econ. Stat., Forthcoming.

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. J. Health Econ. 32 (6), 1263–1277.

McGuire, T.G., Newhouse, J.P., Normand, S.-L., Shi, J., Zuvekas, S., 2014. Assessing incentives for service-level selection in private health insurance exchanges. J. Health Econ. 35, 47–63.

Newhouse, J.P., Manning, W.G., Keeler, E.B., Sloss, E.M., 1989. Adjusting capitation rates using objective health measures and prior utilization. Health Care Financ. Rev. 15 (1), 39–54.

Newhouse, J.P., McWilliams, J.M., Price, M., Huang, J., Fireman, B., Hsu, J., 2013. Do medicare advantage plans select enrollees in higher margin clinical categories. J. Health Econ. 32 (6), 1278–1288.

Newhouse, J.P., Price, M., McWilliams, J.M., Hsu, J., McGuire, T.G., 2015. How much favorable selection is left in medicare advantage? Am. J. Health Econ. 1 (1), 1–26.

Newhouse, J.P., 2017. Risk Adjustment with an Outside Option (working paper).

Panhans, M., 2016. Adverse Selection in ACA Exchange Markets: Evidence from Colorado (working paper).

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

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 International, RTI Project Number 0209853.006, March.

Rothschild, M., Stiglitz, J., 1976. Equilibrium in competitive insurance markets: an essay on the economics of imperfect information. Q. J. Econ. 90 (4), 629–649.

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

Shmueli, A., Messika, D., Zmora, I., Oberman, B., 2010. Health care costs during the last 12 months of life in Israel: estimation and implications for risk adjustment. Int. J. Health Care Finance Econ. 10 (3), 257–273.

Spinnewijn, J., 2017. Heterogeneity, demand for insurance, and adverse selection. Am. Econ. J. Econ. Policy 9 (1), 308–343.

van Kleef, R.C., van Vliet, R.C.J.A., van de Ven, W.P.M.M., 2013. Risk equalization in The Netherlands: an empirical evaluation. Expert Rev. Pharmacoecon. Outcomes Res. 13 (6), 829–839.

van Veen, S.H.C.M., van Kleef, R.C., van de Ven, W.P.M.M., 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. Med. Care Res. Rev., 1–24.

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

Van Barneveld, E., Lamers, L., Van Vliet, R., Vande Ven, W., 2001. Risk sharing as a supplement to imperfect capitation: a tradeoff between selection and efficiency. J. Health Econ. 20 (2), 147–168.

Veiga, A., Weyl, G.E., 2016. Product design in selection markets. Q. J. Econ. 131 (2), 1007–1056.

Zhu, J.M., Layton, T.J., Sinaiko, A.D., 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), 255–275.