Journal of Health Economics 35 (2014) 47–63

Contents lists available at ScienceDirect

Journal of Health Economics

journal homepage: www.elsevier.com/locate/econbase

Assessing incentives for service-level selection in private health insurance exchanges

Thomas G. McGuire[a,b,*], Joseph P. Newhouse[a,b,c,d], Sharon-Lise Normand[a,c], Julie Shi[a], Samuel Zuvekas[e]

[a] Department of Health Care Policy, Harvard Medical School, United States
[b] NBER, United States
[c] Department of Health Policy and Management, Harvard School of Public Health, United States
[d] The Harvard Kennedy School, United States
[e] Agency for Healthcare Research and Quality, United States

ARTICLE INFO

Article history:
Received 27 February 2013
Received in revised form 28 January 2014
Accepted 31 January 2014
Available online 17 February 2014

Keywords:
Health insurance
Exchanges
Adverse selection

ABSTRACT

Even with open enrollment and mandated purchase, incentives created by adverse selection may undermine the efficiency of service offerings by plans in the new health insurance Exchanges created by the Affordable Care Act. Using data on persons likely to participate in Exchanges drawn from five waves of the Medical Expenditure Panel Survey, we measure plan incentives in two ways. First, we construct predictive ratios, improving on current methods by taking into account the role of premiums in financing plans. Second, relying on an explicit model of plan profit maximization, we measure incentives based on the predictability and predictiveness of various medical diagnoses. Among the chronic diseases studied, plans have the greatest incentive to skimp on care for cancer, and mental health and substance abuse.
© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Several provisions of the Patient Protection and Affordable Care Act of 2010 (ACA) are designed to minimize adverse selection in Exchanges (also referred to as Marketplaces).[1] Exchange plans may condition premiums only on age (with restricted rate bands), family size, smoking status, and geography, but not preexisting conditions or other factors. Coverage is regulated. The ACA also mandates that Exchanges engage in risk adjustment and implement temporary risk corridors and reinsurance programs.[2] Risk adjustment is budget neutral: health plans drawing enrollees with lower than average health risk transfer funds to plans with higher than average health risks.[3]

These regulations may not fully address selection problems, however, because Exchange plans may engage in the difficult-to-regulate practice of distorting service offerings to attract "winners" and deter "losers." For example, news stories already contain reports that plans are engaging in aggressive network management, possibly discouraging enrollees requiring more costly treatment.[4] Aggressive network management will also generally lower premiums, making insurance purchase more attractive to good risks.

Assessment of selection incentives is often undertaken by calculating "predictive ratios" for a group with a chronic illness (for example), with the ratio defined as the average risk adjusted payment divided by the average cost for the group (e.g., Pope et al., 2011). One of our contributions is to improve the methodology of predictive ratios. The idea of a predictive ratio is simple: show the

* Corresponding author at: Department of Health Care Policy, Harvard Medical School, United States. Tel.: +1 6174323536.
E-mail address: mcguire@hcp.med.harvard.edu (T.G. McGuire).

[1] On October 1, 2013, U.S. citizens and legal residents who are not eligible for employer-sponsored or public coverage could begin to purchase health insurance through new Exchanges for the January 1, 2014 start date. States can choose to operate their own state-based Exchange, a state-Federal partnership Exchange, or choose instead to rely on the Federal government to perform the function (known as a Federally facilitated Exchange). See Collins and Garber (2013).

[2] The reinsurance and risk corridor programs are to operate from 2014 to 2016 and are intended to create stability in during the transition years (when healthier individuals may delay enrolling). In contrast, risk adjustment is permanent.

[3] These adjustments will take place at the insurance carrier level, based on insurers' aggregate risks across an entire state. Risk adjustment does not apply to self-insured ERISA plans, large group plans, or grandfathered health plans. The federal proposal for risk adjustment is described in DHHS (2013).

[4] Pear (2013) reports on health plans in several states offering low-cost/tight network coverage. One study quoted in the article claims that "...The use of narrow networks may also lead to higher out-of-pocket expenses, especially if a patient has a complex medical problem...".

0167-6296/$ – see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jhealeco.2014.01.009

revenue for a group in relation to the costs for the group. Profitable groups will be attractive to plans, unprofitable groups will be unattractive. While the idea is simple, its implementation in Medicare and in Exchanges has neglected that revenues (in both Medicare and the Exchanges) involve premiums as well as risk adjustment. Premiums themselves involve some "risk adjustment" in that premiums can be up to three times higher for an older than a younger person. In our construction of predictive ratios we anticipate equilibrium premiums to better characterize winning and losing groups.

While predictive ratios are relatively easy to calculate, they are far from a complete description of incentives related to selection in managed care. Managed care plans are usually modeled as making discriminatory decisions about *services* (which is legal though regulated), not about individual persons or groups of people (which is not legal). Thus, a plan might set up a difficult-to-access network of specialists for a disease (e.g., cancer) if it wished to discourage people who would want to use this network in the plan. A plan can do that within limits, but it cannot discriminate on the basis of "pre-existing conditions."

In an alternative to predictive ratios, we use a theory-driven measure to characterize the services a plan would wish, in its own self-interest, to undersupply. Relying on an earlier literature referenced below, we characterize service-level incentives based on an explicit model of plan profit-maximization. A plan will want to stint on quality for services that are predictable by enrollees and predictive of net losses. This second measure, while more precise theoretically, involves more assumptions and empirical work to implement. We must estimate what individuals can predict for various sets of services, and measure the correlation of these predictions with total gains and losses for each person. We show how to implement both measures of incentives based on a "Exchange population" drawn from five panels of the Medical Expenditure Panel Survey (MEPS).

Section 2 contains a brief review of the literature on adverse selection and health insurance markets, emphasizing studies relevant to the new Exchanges. Section 3 presents the economic rationale for our measures of incentives for plans to engage in service-level selection. Section 4 explains how we use the MEPS data to define and construct revenue and cost-related variables used to illustrate our methods. Characterizing plan revenue per person in an Exchange requires us to simulate risk-adjustment. After approximating the risk adjustment to be used in Exchanges, we find the zero-profit plan premiums consistent with the risk adjustment methodology. On the cost side, we assess plan incentives to select across seven disease areas – heart disease, injury, cancer, mental health and substance abuse, lower respiratory, diabetes, and joint and back disorders – a mix of chronic and acute conditions. The measure of predictability requires a statistical model estimating how well individuals can forecast use of various services. Our methods for estimating predictability are described in Section 5. Section 6 presents results for predictive ratios for groups of users, and the measure of incentives to over and underprovide services based on plan profit maximization. Among the disease areas studied, incentives for plans to underprovide services are strongest in the case of cancer, and mental health and substance abuse. A final Section 7 discusses the limitations of our approach, including those related to the uneven rollout of the Exchanges, and some possible next steps for research.

2. Literature review

Enrollees choosing health insurance in their best interest fuels adverse selection: premium differences between plans generally understate cost difference for sicker individuals, making more generous/expensive plans more attractive to sicker types. In an open enrollment environment, inefficiencies arise when plans take actions to discourage the financial losers from joining by limiting coverage for care used by sicker people, or managing certain benefits too tightly. Risk adjustment using demographic variables only partially overcomes these incentives. In the *Handbook of Health Economics* Volume I, Cutler and Zeckhauser (2000, Table 9) summarize thirty studies documenting adverse selection in health insurance. Breyer et al. (2012), in the *Handbook*, Volume II, update the literature review and call attention to potential "indirect selection," which involves plans designing or managing benefits to discourage costly enrollees.

Managed care health plans, when enrolling individuals in products for which they are at risk, are assumed to seek a favorable selection of enrollees by structuring provider networks and managing the administration of benefits. With the notable exception of private health insurance markets operating pre-ACA in the U.S., direct selection (aka underwriting) is prohibited in virtually all markets for health insurance featuring individual choice, including the Netherlands, Germany, Switzerland, major U.S. payers such as private employers, Medicare, Medicaid, and now, in the Exchanges.

Our work is set in the context of individuals choosing among at-risk insurers in an Exchange, as well as individuals insured by Medicare who can choose a product from an at-risk private Medicare Advantage plan. It does not apply to individuals who obtain insurance through self-insured employers or through employers who do not provide a choice among competing, at-risk insurers. Indeed, much of the economics literature fails to distinguish health insurance markets where different insurers compete for individual business and markets with self-insured employers, which historically have comprised about half the American market. Employers offering products from competing insurers will often specify the cost sharing and benefit coverage they want. They cannot control a plan's network or formulary (though they may seek to regulate it) or the plan's utilization management techniques, but can choose not to offer a plan with a network they deem inadequate or management deemed too strict. In short, employers generally control entry into the market for their employees, whereas the literature on selection typically proceeds on an assumption of free entry, as do we. In our context, Exchanges regulate insurance products; some control entry and some do not, but to simplify the exposition we will proceed on the assumption that the Exchange does not control entry.

In competitive individual insurance markets with open enrollment, Rothschild and Stiglitz (1976), assuming Nash behavior, showed that insurers will break a pooling equilibrium by offering less coverage for a lower premium to attract good risks. Glazer and McGuire (2000) applied the same logic to managed care plans skimping on services: there is an incentive to cut back services that are relatively more attractive to higher risk types and offer more generous services that appeal to lower risk types. More precisely, profit maximization implies that plans that are at risk for medical costs have incentives to tightly ration services that are *predictable* and *predictive*. 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. Predictiveness refers to the contemporaneous correlation of use of one service with net revenues per person and governs whether selective rationing will be strict or loose. Services that are both predictable and predictive are especially vulnerable to strict rationing.

Ellis and McGuire (2007) measure predictability, predictiveness, and the consequent incentives to ration services among plans competing in Medicare using data from traditional Medicare (not the managed care component for which data were not available). 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.[5] 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 services (in relation to the average) just as the selection index predicts. This pattern of spending is not observed among enrollees in unmanaged plans. An alternative interpretation, however, is that HMO plans are better at managing diseases that tend to be predictable, i.e., chronic illnesses where the ability to manage care is more feasible, and so reduce spending more for these diseases than for others in relation to less-managed plans.

This latter interpretation is supported by the findings of Newhouse et al. (2013). They find substantial differences in the profitability of various diseases to plans in Medicare; the more profitable diseases, however, are those that are capable of medical management and where insurers face less market power from providers. Importantly for this paper, Newhouse et al., find no evidence of selection across the diseases despite the apparent incentives to do so. We return to these findings in Section 7.

Premium regulation can also cause adverse selection (Pauly, 1985). Community rating restricts premiums to be uniform regardless of expected costs. With a community-rated single premium for all potential enrollees, even if plan design is fixed and underwriting is prohibited, sicker enrollees will be more willing to pay higher premiums for better coverage, and the match between enrollee preferences and plan characteristics will be inefficient. Indeed, community rating and open enrollment are the main ingredients for the notorious health insurance "death spiral" (Cutler and Reber, 1998).[6] Selection stemming from premium regulation is also relevant to Exchanges.[7]

The Federal Employees Health Benefit Program (FEHBP) has run a regulated health insurance market, in essence, an Exchange, for federal employees (including retirees) and their families since the 1960s. The FEHBP, relative to most employers, offers many more plans and is essentially passive with respect to entry. During the early years of the FEHBP, mental health care was covered equally with other care in national plans, but generous coverage proved unviable with individual choice of coverage. Padgett et al. (1993) found strong evidence for adverse selection across plans,[8] explaining plans' cutbacks in coverage and the near death spiral experience for mental health coverage, in spite of Office of Personnel Management (OPM) resistance to cutbacks (Foote and Jones, 1999). In 1980, behavioral health services accounted for 7.8% of total claims costs; by 1997, this had dropped to 1.9%.[9] Exchange plans are not free to reduce nominal coverage in the same way FEHBP plans could do, but selection on networks and management could still materialize.

Some previous state-level Exchanges collapsed, (i.e., they have greatly reduced the coverage options offered or have ceased operations entirely), in large part due to adverse selection (Blumberg and Pollitz, 2009; Wicks and Hall, 2000). In the California Health Insurance Purchasing Cooperative in the 1990s, a voluntary Exchange open to small groups, adverse selection sent the more generous PPO plans into a "death spiral" (Wicks and Hall, 2000). Some early experience from state-level health reform in Massachusetts indicates that individual health insurance markets were subject to adverse selection. Chandra et al. (2011) studied claims costs and prevalence of a chronic illness among those enrolling in the state market pre and post the individual mandate (which lagged by a year creation of highly subsidized plans). Early 2007 saw a spike in enrollment spurred by the mandate, and a shift toward more healthy enrollees. An effective individual mandate deals with selection in and out of health insurance, but not across plans or plan types.

Research groups have simulated how well risk adjustment is likely to ameliorate selection incentives in Exchange plans. Weiner et al. (2012) and Barry et al. (2012) used claims data from 2006 to 2007 from public and private plans to compare average payments and average costs for simulated Exchange plans drawing lower to higher shares of persons with chronic illnesses.[10] At the plan level, risk adjusted revenues tracked average costs well.[11] These papers did not study subgroups or incentives for indirect selection.[12]

In sum, regulation limits what a plan can do to select risks in an Exchange, but a plan can still seek a favorable selection of enrollees by the management of mandated benefits, especially if the Exchange is passive with respect to entry. This general statement is not very helpful, however, in anticipating the functioning of Exchanges. More helpful would be to know: which services are most vulnerable to selection incentives in Exchanges? We develop and implement a method to identify the services most vulnerable.

3. Measuring incentives for service-level selection

We measure the incentives plans have to engage in service-level selection in two ways. The first is the "predictive ratio" for specified groups commonly applied in evaluation of risk adjustment,

[5] In this case the employer, the Group Insurance Commission of Massachusetts, was passive with respect to entry.

[6] In a recent study, Cutler et al (2010, p. 828) find "clear evidence of adverse selection" in the pattern of switching between HMO and FFS plans with competing at-risk plans (in fact the same employment group as Eggleston and Bir (2009)), with higher spending enrollees in an HMO more likely to switch to FFS than lower spending enrollees, and lower spending enrollees in FFS more likely to switch to an HMO than higher spending enrollees. This is likely due to the community-rated feature of premiums in these plans. Einav and Finkelstein (2011) explain this mechanism diagrammatically.

[7] Premiums not reflecting expected costs of groups can also come about from information asymmetries – enrollees might know more about their expected costs than the plans do.

[8] Newhouse (1993) used the FEHBP plans to show how selection can infect estimates of demand response. Empirical estimates of "demand elasticity" for mental health services among FEHBP plans were ten times as high as were found in the RAND Health Insurance Experiment – due to the alert selection behavior by federal employees (Newhouse, 1993).

[9] During this period most private employers were going in the opposite direction, improving coverage for mental health care, and surpassing coverage in the FEHBP. Parity for mental health and substance abuse coverage was implemented in the FEHBP plans in 2001. An evaluation of this benefit expansion confirmed the general finding from earlier research that parity for mental health benefits can be implemented at little increase in total (plan plus OOP) cost in the presence (or with the addition of) managed care. The finding that parity is cost-neutral is surprising until one realizes that managed care plans keep costs down by other means, and they tend to use management more aggressively in the presence of parity (Barry and Ridgely, 2008). Regulation of coverage is not complete protection against service-level selection in managed care.

[10] Weiner et al. (2012) studied chronic illnesses overall and Barry et al. (2012) focused on mental illnesses.

[11] Risk adjustment used was the Adjusted Clinical Groups System, version 9.0.

[12] Predictive ratios for subgroups, such as persons with chronic illnesses, were not reported. Revenue to Exchange plans will depend on premiums as well as risk adjustment methodology.

adapted here to recognize premiums as well as risk adjustment. The second is the predictability–predictiveness measure derived from plan profit maximization.

3.1. Group predictive ratios

In the risk-adjustment literature, a predictive ratio is the ratio of the mean risk-adjusted payment to the mean total cost for a subgroup of enrollees (e.g., those with a particular chronic condition).[13] If this ratio is less than one, risk adjustment systematically underpays for the group.

The typical predictive ratio analysis risk adjusts total costs and assumes that plan revenues equal costs after risk adjustment. This methodology is correct only if the plan is paid exclusively through risk-adjusted payments. The methodology is a bit off in Medicare where enrollee premiums contribute to revenue, and is far off in Exchanges where plans must be self-financing and rules for premiums matter for plan revenues. For use in an Exchange context, we therefore modify the usual predictive ratio to recognize that enrollee premiums and risk adjustment both play into revenue. The numerator of our predictive ratios is the average total revenue a plan receives for members of the group net of risk adjustment. We ignore administrative cost at the plan (assuming in effect administrative costs are proportional to plan spending for the disease). Predictive ratios and our other measure of incentives will be unaffected by factoring costs and revenues down to Exchange plan actuarial values.[14]

3.2. Plan profit maximization and service-level selection

Our second measure of incentives for service-level selection derives from plan profit maximization. We first describe profit, then profit maximization, and finally, the implied measure of incentives.[15] A potential enrollee is indexed by i.[16] Profits are revenue less costs. Revenue from person i, rev_i, depends on the premium the plan charges and the methodology for risk adjustment applied in an Exchange.[17] Next consider costs. A health plan provides services (heart care, mental health care, others) indexed by s, and rations by setting a "shadow price" for each service which can be interpreted as a threshold of clinical need or benefit an enrollee must exceed to receive services. A higher shadow price corresponds to tighter rationing. Let $q = \{q_s\}$ be a vector of shadow prices chosen by the plan to ration these services, and $m_i(q) = \{m_{is}(q_s)\}$ be the vector of spending on service s enrollee i receives as a function of their own characteristics and the shadow price. We assume rationing can be done without incurring direct resource costs.[18]

The level of spending enrollee i receives for service s, $m_{is}(q_s)$, is that which equates the marginal benefit of spending for that enrollee equal to the shadow price q_s. Let c_s be the share of spending paid by plan enrollees for service s, and the plan pays a share of $(1 - c_s)$.[19] We can now write the expression for profit for enrollee i as $\text{rev}_i - \sum_s m_{is}(q_s(1 - c_s))$. Given enrollment in a plan, the plan increases profits by rationing more tightly, tending to increase each q_s.

The countervailing incentive is created by the plan's interest in attracting and maintaining membership. Less strict rationing (lower q's) attracts more enrollees.[20] Individuals enroll in a plan on the basis of what they *expect* to receive in the plan. Let $\hat{m}_{is}(q_s)$ be the services enrollee i expects to receive in a plan rationing by the vector q. From the standpoint of the plan, individual i enrolls in the plan with a probability $n_i(m_{is}(q_s))$ a function of the vector of shadow prices.[21] We can now write the complete expression for profit at the plan, $\pi(q)$, as depending on who joins and the profits per enrollee. Both probability of joining and profits per enrollee are a function of the shadow price rationing:

$$\pi(q) = \sum_i n_i(\hat{m}_{is}(q_s)) \left[\text{rev}_i - \sum_s m_{is}(q_s)(1 - c_s) \right] \tag{1}$$

The plan maximizes (1) with respect to each q_s; the solution characterizes plan rationing to maximize profit. Presence of the $\hat{m}_{is}$ terms implies that enrollee expectations or predictions will matter to profits. Intuitively, unless service use can be predicted by enrollees, setting a shadow price low or high will fail to attract/deter certain types of enrollees. In the extreme, enrollees could not anticipate at all whether they were likely to be low or high users, and thus everyone expected themselves to be average. In that case, setting a shadow price low or high would attract/deter everyone equally and thus be no use as a selection device. In the other extreme in which everyone can perfectly predict what they will use, setting the shadow price for that service will be very effective in achieving selection on that service. "Predictability" of a service thus determines the power of service as a selection device.

What determines whether a plan would want to use predictable services to attract or deter users? The profit expression (1) shows that a correlation between expected use of a service, $\hat{m}_{is}(q)$ and profit, $\text{rev}_i - \sum_s m_{is}(q_s(1 - c_s))$ indicates whether a service tends to be used by the winners or losers. Within a population, when (expected) use of a service s is positively correlated with profits, the plan will want to ration loosely (set q_s low) to attract those types. When use of a service is negatively correlated, the plan will want to ration tightly (set q_s high). The stronger is the negative

[13] Predictive ratios are the primary basis of the recent evaluation of the CMS-HCC model of risk adjustment used to pay private health plans in Medicare. See Pope et al. (2011). The General Accounting Office used average profit/loss by subgroups to evaluate risk adjustment in Medicare for disabling chronic conditions (GAO, 2011). Barry et al. (2012) and Weiner et al. (2012) construct predictive ratios at the plan level according to the degree of adverse selection of risks drawn by the plan.

[14] A qualification is necessary here: Plan coverage will be non-linear, including deductibles and OOP maximum. To the degree that different types of services tend to fall in these coverage ranges, plan costs will not be simply proportional to actuarial value by service.

[15] Our measure is based on a characterization of health plan profit maximization. Health plan behavior also depends on conditions imposed by market equilibrium. In any equilibrium, profit maximization will have to be satisfied. Equilibrium conditions play in here by their effect on premiums described below. In a symmetric equilibrium, all plans would be following the same profit maximization behavior described here. Rothschild–Stiglitz type models focus on equilibrium conditions and imply service-level rationing.

[16] We present this as an individual health insurance market. We discuss how we handle families in the next section.

[17] We do not need the mechanics of Exchange accounting yet. This comes in the next section when we describe how we measure revenue per person.

[18] Adding a cost factor could change our results if the cost of rationing differed by service type, but we have no basis for measuring such a cost.

[19] We assume cost sharing is fixed. Even if all services are covered with the same benefit, the plan and enrollee share of different services could differ, for example, if some services were more likely to fall within deductible limits and have a higher share of enrollee payment.

[20] This ignores the possibility that better coordinated care reduces services and improves outcomes. We assume that more care is better from the standpoint of the patient.

[21] The premium the individual must pay is regarded as fixed. We can also ignore the cost sharing so long as the shadow price is above the cost sharing (the real price) the patient must pay, more services will make the plan more attractive to the enrollee. Cost sharing tends to be low in managed care, and efficient shadow prices are one.

correlation, the more a service is used by financial losers. This is what we mean by "predictiveness." A service predictive of profits should be rationed loosely, and a service predictive of losses should be rationed tightly.

Ellis and McGuire (2007) use conditions for maximization of (1) to derive an expression for how predictability and predictiveness together determine incentives.[22] They show that plan incentives to ration tightly at the service level are proportional to the product of measures of demand elasticity, predictability and predictiveness. Specifically, letting I_s be an index of the incentives to ration:

$$I_s \propto -\varepsilon_s \cdot cv_s \cdot \rho_s \tag{2}$$

where ε_s is demand elasticity (a negative number), cv_s is the coefficient of variation of predicted spending on service s, and ρ_s is the correlation between expected spending on service s and plan gain/loss.[23]

Demand elasticity, which we do not estimate here, scales the effect of predictability and predictiveness. We do estimate the other two components of (2). The coefficient of variation of predicted spending has a natural interpretation in terms of predictability of a service. The lower limit of predictability is when everyone expects to be average, in which case the cv_s is zero. As individuals' abilities to forecast their difference from the average improves, the cv_s goes up. To compute cv_s for each service we fit a model of predicted spending, and calculate the *cv* of the predictions. Note that the coefficient of variation is standard deviation divided by mean spending, scaling all services on a comparable (unit-free) measure.

The correlation between individuals' expected spending on a service s and gain/loss represents the *predictiveness* of service s,[24] a correlation which could be positive or negative, making I_s itself positive or negative. Large positive values characterize services that are both predictable and whose use predicts gains. The plan will want to ration these loosely. I_s will be near zero for services that are either not predictable, or if they are predictable, do not correlate with winners or losers. The plan has no incentive to use these services for selection. Large negative values characterize services that are predictable and whose use predicts losses. The plan will want to ration these services tightly.

4. Empirical application

4.1. Data and the exchange population

We use data from the Medical Expenditure Panel Survey (MEPS), a nationally representative survey of the civilian non-institutionalized U.S. population conducted annually since 1996. Each year MEPS collects information on approximately 33,000 individuals, enlisting a new panel of respondents that it follows for two years. Data are collected in five rounds of interviews covering the two-year period. The Household Component (HC) is the source for personal and household characteristics, including insurance coverage and self-reported health and health conditions. The HC is also the source of data on medical "events" (e.g., an inpatient stay or office visit) including information about diagnoses, procedures, and payments from various sources. The HC data are supplemented with information from the Medical Provider Component (MPC), based on phone surveys of hospitals, physician offices, pharmacies, and home health agencies. We use Panels 9 (2004/05) through 13 (2008/09), requiring participation in both years of the panel (dropping those who die during their first survey year or otherwise leave the sample). We take advantage of the two-year panel structure of MEPS to implement risk adjustment; our first year of study for expenditures will be 2005, because we do not have data on individuals' medical events (analogous to claims) for 2003 and so cannot predict spending in 2004.

We select a population of individuals and families who would be eligible to enroll in state-level Exchanges under current law based on their income, insurance, and employment status. We identify adult, non-elderly individuals (aged 18–64) in households earning at least 138% of the federal poverty level and children in households with income of at least 205% of the federal poverty level.[25] We select for the Exchange population those who live in a household where an adult is: ever uninsured, a holder of a non-group insurance policy, self-employed, employed by a small employer, or paying an out-of-pocket premium for their employer-sponsored health insurance (ESI) plan that is deemed to be unaffordable (as defined in the ACA).[26] In total, we have 20,865 individuals from MEPS, each with two years of data.[27] We do not take into account less than full compliance with the mandate, as for example would be the case if younger groups required to join were nonetheless not enrolling. In essence, we seek to model incentives once the Exchanges are up and running as designed.

Table 1 summarizes some statistics on our sample. The population contains a relatively high proportion of Hispanics, lives disproportionately in the South, and exhibits a large range of income, with a third of the sample having incomes over 400% of the poverty line. We compare the self-rated health and mental status of our adult Exchange population to the health status of those with Employer-Sponsored Insurance (ESI) (Table 2). For each Exchange observation, we randomly draw one observation from the ESI population (with replacement) from the same five-year age band, sex, region and MEPS panel. After this simple matching procedure, Table 2 shows that the overall health and mental health status of the Exchange population is slightly worse than the ESI population.

MEPS respondents tend to underreport ambulatory visits, though other types of care are generally reported accurately (Hill et al., 2012). Because MEPS is a community sample, some individuals in long-term care are underrepresented. AHRQ staff have developed corrections for these biases which we implement here (Zuvekas and Olin, 2009). The major limitation of using the MEPS data for our purposes is that observed health care expenditures are affected by the household's insurance coverage. Because we wish

[22] The form of the index follows from the derivative of profit in (1) with respect to q_s, normalized by mean spending on each service.

[23] See Ellis and McGuire (2007) for derivation of (2). The full expression is $I_s \propto -\varepsilon_s \cdot cv_s \cdot (\rho_s - C)$ where C is a constant. We ignore the constant in this analysis.

[24] To be fully accurate, in all cases it is the plan's predictions that matter, because it is the plan's behavior that is of interest. Thus, with respect to service spending, it is the plan's beliefs about individual predictions that matter, and with respect to total spending governing gains and losses, it is the plan's expectations about total spending. We do not make an empirical distinction between individual and plan predictions of spending and just refer, in this paper, to predicted spending.

[25] Annual household income from each year is inflated to 2009 dollars using the Consumer Price Index (CPI-U) published by the Bureau of Labor Statistics, and we apply 2009 federal poverty guidelines for the 48 contiguous states available online at: http://aspe.hhs.gov/poverty/09poverty.shtml. We follow the methodology of the Kaiser Family Foundation that uses these income criteria to select the population eligible to purchase insurance through an Exchange (Trish et al., 2011). Adults and children in households with lower incomes are deemed to qualify for Medicaid. We do not simulate employer behavior as does the CBO model (CBO, 2011).

[26] Small employers are either (1) those with fewer than 50 employees or (2) those with fewer than 100 employees and who report only one business location. The ACA states that individuals whose out-of-pocket premiums for employer-sponsored insurance exceed 9.5% of family income will be eligible to purchase health insurance through an Exchange.

[27] We used no additional weighting of observations.

Table 1
Descriptive statistics of exchange population, MEPS 2005–2009, $N = 20{,}865$.

Data reported as percentages, unless noted	
Male	50.8%
Age	
0–18	13.5%
19–64	86.5%
Race	
Non-hispanic white	51.1%
Non-hispanic black	12.5%
Hispanic	28.8%
Asian	5.2%
Other	2.4%
Metropolitan area	85.0%
Region	
Northeast	13.9%
Midwest	19.1%
South	38.7%
West	28.3%
Marital status	
Married	51.2%
Widowed, divorced, separated	12.5%
Never married	27.9%
Inapplicable	8.3%
Average family size (*n* persons)	2.8
Education level	
Less than high school degree	19.4%
High school degree	29.5%
Some college	14.7%
College degree or more	25.7%
Inapplicable[a]	10.8%
Mean individual income ($2009) [Standard Deviation]	$33,300 [$31,500]
Poverty status (based on family income)	
<138% FPL	2.8%
139–200% FPL	18.2%
201–300% FPL	28.0%
301–400% FPL	17.5%
400% FPL or higher	33.5%
Employment status	
Continuously employed	70.5%
Continuously unemployed	10.2%
Both employed/unemployed	9.2%
Inapplicable[b]	10.1%
Self-reported health status	
Excellent	30.6%
Very Good	33.3%
Good	27.4%
Fair	7.3%
Poor	1.4%
Self-reported mental health status	
Excellent	41.5%
Very Good	30.5%
Good	23.8%
Fair	3.7%
Poor	0.5%

[a] Most observations in "Inapplicable" are children with very young age. It also includes "not ascertained", "don't know" and "refused".
[b] Most observations in "Inapplicable" are people not in labor force. It also includes "not ascertained", "don't know" and "refused".

to assess plan incentives in an Exchange, we would ideally like to observe spending conditional on the insurance the household would buy in the Exchange rather than their actual coverage, which could be none. This limitation is endemic to simulation research on the ACA, and indeed in the empirical risk adjustment field generally, where models are often fit on "outside" data.[28]

[28] For example, risk adjustment models used to pay managed care plans in Medicare are fit on data from traditional Medicare.

Table 2
Comparing health and mental health status of adults in exchange and employer-sponsored insurance populations, $N = 18{,}047$.[a]

	Mktplace	ESI	*p* Value for difference[b]
Health status, %			
Excellent	27.3	30.6	<0.001
Very Good	33.8	35.5	0.001
Good	29.1	26.0	<0.001
Fair	8.2	6.6	<0.001
Poor	1.7	1.3	0.015
Mental health status, %			
Excellent	39.5	43.9	<0.001
Very Good	30.8	31.8	0.047
Good	25.1	20.4	<0.001
Fair	4.0	3.4	0.005
Poor	0.6	0.4	0.007

[a] Only adults (age 19–64) are included in the comparison, so sample size is less than the total sample size of 20,865.
[b] Two-tailed *Z*-tests with null hypothesis that proportions p1 = p2 in each category. Each Exchange observation was matched with an Employer Sponsored Insurance (ESI) observation by age, sex and region. Matching was done with replacement among the ESI sample.

4.2. Plan revenues

To measure incentives, we first construct revenue received by health plans from each person, taking account of some features of Exchange plan payment. In particular, payments to plans are subject to risk adjustment. Plans will thus set premiums in a market in a way that accounts for the risk adjustment scheme and whatever regulations the Exchange imposes.

We risk adjust with the Hierarchical Condition Category (HCC) model, a version of which was recently chosen as the basis for federally facilitated Exchanges (DHHS, 2013). An analogous version of this model is used by Medicare to pay Medicare Advantage plans. That model, the CMS-HCC model, uses individual demographics and indicators of major medical conditions in a base year to predict an individual's health care expenditure for the next year. It maps individual diagnoses from ICD-9 codes into one of 70 hierarchical condition categories (HCCs) to predict costs. Diseases within an HCC are similar clinically. Each individual is given a (0, 1) indicator for each HCC, and these become part of a linear regression model predicting cost. The coefficients from this model are the "weights" on age, sex, HCC and other factors used in risk adjustment (Pope et al., 2011). We use the same age categories as the CMS-HCC model.

The model proposed for federal Exchanges is more complex. It uses 100 HCCs, has more interactions, is concurrent rather than prospective, and is estimated separately for children and adults, and separately for each of the plan actuarial values in the Exchange.[29]

Our risk adjustment model diverges from the Medicare CMS-HCC model in a several ways to accommodate the MEPS Exchange population and rules. First, our risk adjustment model excludes variables indicating Medicaid and disability status because these are not applicable to the population that will be insured through the new Exchanges. Second, whereas the CMS-HCC model uses 5-digit ICD-9 diagnosis code to classify diagnoses, the MEPS public use files do not include 5-digit ICD-9 codes. We use the 3-digit

[29] Data for these estimations is from Truven MarketScan data from private health plans and corporations. The model for the federally facilitated Exchanges uses current-year experience to risk adjust partly because, for many Exchange participants, there will be no "prior year" data available for risk adjustment. Here, since we have a prior year for all observations, we can use prospective risk adjustment.

ICD-9 codes, which are publicly available.[30] Documentation of the CMS-HCC model indicates that moving from 3 to 5-digit classification does little to improve model fit in MEPS.[31] Moreover, in MEPS, diagnostic data come from household reports which lack the specificity and precision of physician reports (AHRQ, 2011). Third, we do not include the full set of 70 HCC indicators because of limitations of our sample size, nor do we include interaction effects for the same reason. We used the 38 HCCs with more than 20 observations. The remaining 32 HCCs are aggregated into one of two categories based on the average annual health expenditure of individuals in the HCC: high if average expenditure is larger than $10,000 and low otherwise. Dummy variables indicating these expenditure categorizations are included in our model. Finally, we limit plan financial responsibility to the first $50,000 of spending for each enrollee per year, reflecting mandatory reinsurance in the first two years of Exchanges, and the possible continuation of some form of voluntary or mandatory reinsurance later on.[32] Table A.1 in the Appendix lists average spending per person by the HCCs that we include in our risk adjustment model. Notably, 80% of the Exchange population has no HCC in the prior year. For such a person, the risk score will depend only on the relative weights for age and sex. We estimate risk adjustment weights using the same method as the CMS-HCC model, fitting an OLS regression on the set of (0, 1) variables included as risk adjustors (Pope et al., 2011). We will refer to the risk adjustment system we use as CMS-HCCs, even though we modified it in the ways just described.

In addition to the CMS-HCC model, we also estimate a model with just age and sex categories from the CMS-HCC model. This is useful to compare with the model with diagnoses and to include those enrollees in an Exchange who have no medical history with which to construct HCCs. The age-sex model corresponds to the "new enrollee" model used by Medicare to pay for beneficiaries turning 65. Estimates from both risk-adjustment models are in Appendix Table A.2. The adjusted R^2 for the age-sex model is 0.186 and for the CMS-HCC model 0.272, higher than typically found, largely because we trim top-end expenses.[33]

The ACA specified that premiums are to be based on age (with regulated rate bands), smoking status, geography, and family size, but not on pre-existing conditions, sex, or other factors. We set four age categories: child (0–18), young adult (19–34), middle-age adult (35–54) and older adult (55–64). To mimic state geographic divisions with the national MEPS data, we use the four census regions, which, together with the four age categories, give us a total of 16 (4×4) premium categories. We do not use smoking status because this variable, conditional on age, is weakly associated with health care costs in the Exchange population.[34] For purposes of this analysis, we treat families as a collection of individuals, with each family member contributing a premium to plan revenues.[35]

Integrating premiums, which themselves partially "risk adjust" payments to plans (e.g., those plans with higher expected risk-adjusted enrollee cost will quote higher premiums), with a risk adjustment methodology presents a new set of questions to regulators and researchers. CMS actuaries developed a formula determining risk-adjustment transfers that attempts to net out the "risk adjustment" accomplished by premiums.[36] We approach the matter differently and take account of market factors in premium determination.

Specifically, we calculate premiums for each category (e.g., older adult, urban northeast) as the premium necessary to just cover costs if a plan draws a random sample of persons in that category.[37] The risk adjustment payments net to zero across plans; thus, plans with sicker (higher use) than average enrollees will receive payments financed by the remaining plans. With the risk-adjustment methodology specified we can solve for the premiums that equalize average plan revenue to average plan cost within each premium category. We calculate premiums for each category as the premium necessary to just cover costs if a plan draws a random sample of persons in that category.[38] The formula for premiums for each premium category is in the Appendix. Our method assumes one risk adjustment model applies to all plan levels (Silver, Gold, etc.).

Appendix formula (A.2) for premiums covers cases in which all medical spending is adjusted, none is risk adjusted, or any share in-between. Specifically, our Exchange risk adjusts a share, σ, $0 < \sigma \leq 1$, of the average cost of all enrollees. Choice of σ involves an under-appreciated tradeoff in Exchange payment design. Suppose $\sigma = 1$ and the Regulator risk adjusts all payments. Because risk adjustment categories are correlated with premium categories (age is in both the CMS-HCC model and premiums, and HCCs are correlated with age), risk adjusting all costs will minimize the residual of costs to be picked up by premium categories. The upside of risk adjusting all costs is that the overall fit of the payment system is maximized. The downside is that competition will "compress" premiums toward the population mean costs. When premium categories are closely correlated with risk adjustment categories, setting $\sigma = 1$ means we move as close as possible to "community rating" of premiums. The young will be overcharged and the old will be undercharged for their health insurance, with the unintended consequences of some young healthy people potentially dropping out of the pool (mandates and penalties notwithstanding) and older people "overinsuring" by buying too much health insurance.

In the current paper we first work through payment system design and incentives with a value of $\sigma = 0.5$, sacrificing some

[30] In our model we assume that the 3-digit code we observe in the data corresponds to the smallest ICD-9 code that starts with those three digits. For example, an ICD-9 code of 003 in MEPS is assumed to represent 0031, which is the smallest code within the 003 category.

[31] MEPS documentation states: "DxCG Inc. staff have examined how using 3-digit diagnoses (rather than 5-digit codes) would affect the prospective DCG/HCC model's performance. They concluded that, although using 3-digit codes would reduce the model's specificity in clinical classification and its predictive accuracy, the loss in specificity and predictive power was small." (AHRQ, 2008, p. C-2).

[32] $50,000 is the "attachment point" (where reinsurance kicks in) used in Winkelman et al. (2011) in reviewing reinsurance rules for Exchange plans. When a person's spending exceeds $50,000 in a year, we factor down every spending event evenly to cap annual spending at $50,000. In this way we keep all events for purposes of classifying a person for risk adjustment or into spending categories as explained below. We ignore the risk corridors, which are transitory.

[33] We fit an earlier model without the trim at $50,000 and the CMS HCC model had an R-squared of 0.16, only slightly higher than typical.

[34] We found this in our earlier paper on premiums and risk adjustment with these data (McGuire et al., 2013). Smoking status is also problematic from the standpoint of accurate reporting.

[35] Individuals are also treated as independent observations in our empirical analyses.

[36] The complex formula is described in the Federal Register 77 FR 73141 in which "allowable rating factors" refer to premiums (DHHS, 2012).

[37] Exchange operations are financed by taxes on plans. We ignore any such taxes and assume that the risk-adjustment scheme generates no net revenue for the Exchange. Premiums and risk adjustment weights can be solved for simultaneously to find the best-fitting payment system. When plans will be setting premiums, deriving the risk adjustment weights that lead to the best fit of payments to costs generally requires "back solving" for the risk adjustment weights. In an earlier paper, we showed how this can be achieved with constrained least squares regression (McGuire et al., 2013).

[38] An alternative approach would be to assume a fixed medical loss ratio, such as the 85% minimum for the ACA, which is similar to assuming a fixed markup due to monopoly power. Ericson and Starc (2012) studied pricing in the individual health insurance market created in Massachusetts and found some evidence for differential markups by age (older people may be less demand responsive). If this were so, this might alter the incentives to ration across disease areas.

Table 3
Number of patients and year 2 expenditures by CCS group, exchange population in MEPS 2005–2009.

	N	*N* (%)	CCS expenditure	Total expenditure		
				Mean	Exp. (%)	Std. Dev.
1. Heart Disease	300	1.4%	$5009	$11,847	8.1%	$13,882
2. Injury	959	4.6%	$2690	$6142	13.4%	$9192
3. Cancer	239	1.1%	$6867	$12,564	6.9%	$15,248
4. Mental Health and Substance Abuse	573	2.7%	$1979	$6121	8.0%	$7742
5. Lower Respiratory Disorders	379	1.8%	$2379	$7131	6.2%	$9654
6. Diabetes	539	2.6%	$2342	$8698	10.7%	$11,515
7. Non-Traumatic Joint and Back Disorders	1078	5.2%	$2226	$6940	17.1%	$9802
Total (includes other categories)	20,865	100.0%		$2100	100.0%	$5245

CCS groups for each disease group are as follows: Heart Disease (96, 97, 100–108), Injury (225–236, 239, 240, 244), Cancer (11–45), Mental Health and Substance Abuse (650–663), Lower Respiratory Disorders (112, 27–133), Diabetes (49, 50), Non-Traumatic Joint and Back Disorders (201–205).

payment system fit but avoiding too much premium compression, and then compare this, in Section 6.3 below to a value of 0.8, with better fit but more premium compression. Table A.3 in the Appendix contains the premiums for each premium group consistent with zero profit for each premium category.[39] With the fully specified risk adjustment and premium payment system we have defined revenues for plans in Exchanges that when joined with costs determine incentives.[40]

4.3. Medical spending and services

We define service categories based on the AHRQ clinical classification software (CCS) which groups the approximately 12,000 ICD-9-CM codes into 260 mutually exclusive categories that are clinically meaningful. AHRQ researchers further grouped the 260 CCS codes into 23 groups based on sample size and magnitude of expenditures (Machlin et al., 2009). Because 23 groups are still too many for our purposes, we combined some conditions from Machlin et al. (2009) into the seven relatively common conditions, shown in Table 3.[41] Four of these conditions are chronic illnesses – heart disease, cancer, mental illness and diabetes – and seem good candidates for selection-related incentives.

MEPS events contain up to three or four CCS codes (depending on the type of file), corresponding to primary, secondary, and tertiary diagnoses. Overall, in our data 13% of events contain more than one CCS code. We classify events according to the first-listed CCS code, and then, if they are not within one of our identified CCS categories, we classify them into one of the categories based on the second-listed code. This happens slightly more than one percent of the time. Moving on to the third or fourth-listed code is rare. To be placed in one of the disease categories in a year we require an enrollee to have two outpatient events with a diagnosis within the CCS group or one inpatient event with the diagnosis.[42]

Table 3 reports the unweighted distribution of people and total expenditures by our seven categories based on the CCS categories for the second year of each of our panels, the years 2005 through 2009. Spending is in 2009 dollars,[43] and has been trimmed to a maximum of $50,000 per person per year as described above. Note that while spending events fall in just one CCS category, a person could have spending in more than one CCS category during a year, so the groups are not mutually exclusive. The table shows, for each group of people, spending in the category of services that define the group (e.g., heart disease) and the total spending for all services for people in that group. The percent column under total expenditures shows the percent of total expenditures accounted for by people that group.

Membership in each group ranges from about one to five percent of the total population, with sample sizes from 239 to 1078. Mean spending within a CCS category is highest for the Heart Disease and Cancer groups, and lowest for Mental Health and Substance Abuse. This pattern of highest and lowest remains true for total expenditures as well. For each group, expenditures within the CCS itself comprise a third to a half of total expenditures. These groups are all three or more times as expensive as the average Exchange participant. They are more expensive not just in the CCS expenditure category but for other services as well. The Cancer group is just over one percent of the population but almost seven percent of total costs. Year 2 spending amounts, summarized in Table 3, will be the basis for calculating plan gains and losses, and form the dependent variable in our predictive modeling.

In sum, we define groups for purposes of constructing predictive ratios using the seven service categories, and compute estimated revenue and plan costs for each group.

5. Modeling predictions

Evaluating the selection index for each service is based on a set of models that seek to represent what potential enrollees are able to predict about their health care spending at the time they decide about health plan membership. This task is distinct from fitting a risk adjustment model based on available data elements

[39] In practice, premiums have to cover administration, marketing and profits, as well as claims costs. In Exchanges, the medical-loss ratio must be at least 80%. As noted above, for purposes of analyzing selection incentives, it is the medical claim cost part of premiums that is relevant, which is displayed in <***>Table C. Recall also that we ignore any administrative costs that may differ by service category.

[40] We can calculate the overall fit of payments to costs as the share of the overall variance in individual-level health care costs explained by the sum of premiums and net risk adjustment payments received by the plan. In the case of the CMS-HCC model, this share, analogous to an R^2 statistic, is 0.252. For the age–sex model, the share is 0.185, very close to the values in <***>Table B that are based on standard risk adjustment.

[41] Six categories – heart disease, injury, cancer, mental health and substance abuse, lower respiratory, and diabetes – we took directly from Machlin et al. (2009). We combined non-traumatic joint disorders and back disorders into one, based on sample size and that these conditions would be treated by the same type of physicians.

[42] Fullerton et al. (2012) use a two-claim method and cite other studies also using this method. We allow one inpatient claim in the group to put a person into the category.

[43] Expenditures are inflated to 2009 dollars using the unadjusted Medical Care component of the Consumer Price Index, published by the Bureau of Labor Statistics.

Table 4
Spending distribution in year 1 by CCS group.

	% of full sample with positive spending	Positive spending[a]		
		Low	Medium	High
1. Heart Disease	1.5%	$392	$1772	$16,679
2. Injury	5.3%	$228	$863	$7542
3. Cancer	1.2%	$277	$1290	$13,941
4. Mental Health and Substance Abuse	3.0%	$350	$1198	$5412
5. Lower Respiratory Disorders	2.0%	$240	$683	$4814
6. Diabetes	2.4%	$482	$1318	$5847
7. Non-Traumatic Joint and Back Disorders	5.1%	$204	$725	$5799
Total (includes other categories)	75.6%	$221	$1062	$8045

[a] The three columns show the mean of year 1's spending by tertile among those with positive spending.

that are not "gameable" and satisfy other criteria relevant for payment.[44]

5.1. Information set: right-hand side variables

Table 4 contains some information about the distribution of Year 1 spending in CCS categories that we use as independent variables in our model. A small share of the population makes it into each CCS category each year. Positive spenders were subset into tertiles and indicators created for low, medium, and high. The likelihood of being in the category in Year 2 conditional on being in it in Year 1 is as follows for our seven categories: Heart Disease, 0.33; Injury, 0.24; Cancer, 0.39; Mental Illness, 0.47; Lower Respiratory, 0.25; Diabetes, 0.66; Joint and Back, 0.42. This is a simple indicator of predictability.[45] Injury has a relative low year-to-year connection whereas the diabetes classification is very likely to persist.[46]

We include the following set of Year 1 variables in our base models that predict Year 2 spending: age–sex combinations from the CMS-HCC model, indicators for category of self-rated health and mental health status, indicators for the tertile of total spending, and indicators for the tertile of spending within the CCS category among those with positive spending. By including a variable we are in effect assuming that the individual knows this information, and furthermore, knows how it relates to expected spending in Year 2. This information set is similar to that used in Ellis et al. (2013) in their two-year data.[47] We employ a parallel specification for each spending category to investigate the relative predictability of spending by CCS group.[48]

One potential right-hand side variable not included in our models is the source of insurance coverage (e.g., employer-based, Medicaid, etc.). In survey data, the partial correlation of plan and spending will reflect both the effect of plan on spending (moral hazard) and individual's choice of plan based on expected health care use (adverse selection). In principle we would want to adjust for the moral hazard effect and use predictions of spending for a standardized health plan (for a private managed care plan of the type expected in an Exchange.) Our previous experience with MEPS[49] suggests that the selection (health status) effect may dominate the moral hazard effect in the data, so we omit this variable from the model.

5.2. Estimation methods

We estimate models for the five MEPS panels combined, using second-year spending as the dependent variable. Each individual appears once in the estimation. Our dependent variable throughout is annual spending, limited to $50,000, in total and partitioned into services. Because of the semi-continuous nature of spending, we estimate two-part quasi-likelihood generalized linear models. In the first part, we fit logistic regression models of the probability of positive spending in Year 2 as a function of the right-hand side variables described above. In the second part, we estimate a quasi-likelihood generalized linear model for those individuals who had positive spending. We use a log link for the mean function and a Poisson function for the variance. The latter choice reflects our assumption that the variance is proportional to the mean.[50] We then determine (unconditional) predicted spending for each individual by combining the estimates from both parts of the model. This is accomplished by multiplying the probability of positive spending from part 1 by the expected spending obtained from part 2 for each individual in the sample.[51] The coefficient of variation of expected spending for each service category, cv_s, is computed from these predictions.

Assessment of the fit of our models is undertaken using a variety of techniques. For part 1, we examine the area under the Receiver Operator Characteristic (ROC) curve. An ROC area of 1 indicates perfect discrimination between those who have positive spending and those who do not while an ROC area of 0.5 indicates poor discrimination, i.e., no better than a coin flip. For part 2, we examine the Akaike Information Criterion (AIC), a relative measure of model complexity and fit. We compared models that included only age-sex terms as right-hand side variables with models that included more right-hand side variables to investigate how much information about prior health status and prior spending contribute to

[44] See Breyer et al. (2012) for a recent discussion of criteria for risk adjustment variables.

[45] These numbers represent the probability of a claim in a second year in the same diagnostic category, as opposed to the likelihood that the illness itself remains. Diabetes, for example, persists with a much higher likelihood higher than 0.66.

[46] Since there is no cure for diabetes, the lack of a claim in Year 2 suggests that the individual did not seek care for the disease.

[47] Ellis et al. (2013) used two years of claims data so they did not have self-rated health and mental health. They did have a much larger sample size and so included measures of spending for all of their 33 service categories in all models.

[48] Papers that have studied what individuals can predict emphasize the importance of the individual error term that can be thought of as being composed of a time-invariant piece, an autoregressive piece, a time-varying piece, and a part that is purely random. Newhouse et al. (1989) uses repeated observations on individuals to incorporate the first two of these terms. With only two years of data we cannot use individual fixed effects.

[49] See Cook et al. (2009a,b).

[50] We also estimated models using a Gamma function for the variance component that reflected the variance is proportional to the mean squared. Results were little affected.

[51] Our right-hand side variables are observed on all individuals, not just those who had positive spending. We thus obtain expected spending for all individuals (not just those who had positive spending) by multiplying the observed right-hand side variables by the estimated regression coefficients and transforming to the dollar scale.

Table 5
Regression results and fit of models for spending prediction.

	C statistics for probability of spending (logistic) model		AIC for conditional spending (quasi-GLM) model			Mean absolute error for unconditional spending predictions, $	
	AS	AS + health status & spending	N	AS	AS + health status & spending	AS	AS + health status & spending
1. Heart Disease	0.790	0.870	300	2,883,848	2,501,567	139.2	130.6
2. Injury	0.591	0.720	959	5,324,811	5,111,488	234.9	223.6
3. Cancer	0.778	0.877	239	3,149,826	2,713,100	153.2	134.1
4. Mental Health and Substance Abuse	0.609	0.857	573	1,376,799	1,041,779	105.4	75.3
5. Lower Respiratory Disorders	0.669	0.789	379	1,516,565	1,275,343	84.3	74.4
6. Diabetes	0.787	0.924	539	1,670,648	1,162,906	113.8	72.3
7. Non-Traumatic Joint and Back Disorders	0.693	0.829	1078	4,914,335	4,405,509	212.9	185.7
Total	0.669	0.811	15,133	79,985,168	69,393,399	2420.0	2128.9

"AS" means the results are from models using age–sex covariates. "AS + health status & spending" means the results are from models using age–sex plus health status and spending covariates. *N* for total (15,133) is the number of individual with positive spending in year 2.

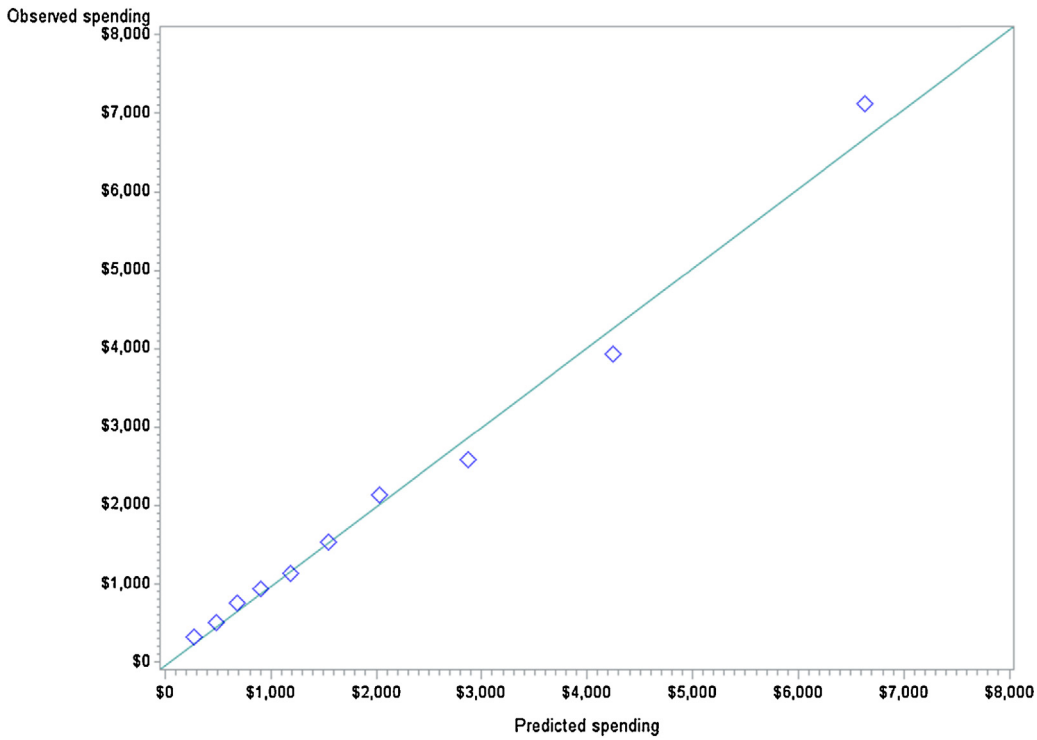

Fig. 1. Observed and predicted total spending by decile of predicted spending.

fit. Because we are primarily interested in how well our model predicts, we compute the mean absolute prediction error of Year 2 spending and we assess how well the observed mean spending agrees with the predicted mean spending. In particular, we graph the mean spending by decile of predicted spending against observed spending for individuals in the predicted decile.

5.3. Results for two-part model

Table 5 summarizes statistics on the fit of our models, for each of parts 1 and 2, and then overall.[52] The two columns in each case compare the fit statistic with the age-sex only model, and then for the model with previous spending and self-assessed health status. The mean absolute error for each spending category, measured in dollars and capturing the overall fit of parts 1 and 2 together, is easiest to interpret. The mean absolute error is influenced by the level of spending which differs across categories. (Our measure of cv_s normalizes by category spending and so is not subject to this problem for comparability.) The last two columns can be used, however, to compare the effect of the inclusion of past spending and health status on the accuracy of predictions by category. The mean absolute error decreases the most in cancer, mental illness, diabetes, and joint and back disorders, indicating the information in past spending and self-assessed health is more predictive for these groups. It is not surprising that the improvement in prediction is less for injuries and lower respiratory disorders. The relatively small improvement in heart disease was unexpected.

Fig. 1 graphs the mean of predicted and actual spending by decile for total spending. All points fall near the 45-degree line indicating

[52] Complete regression results are available from the authors upon request.

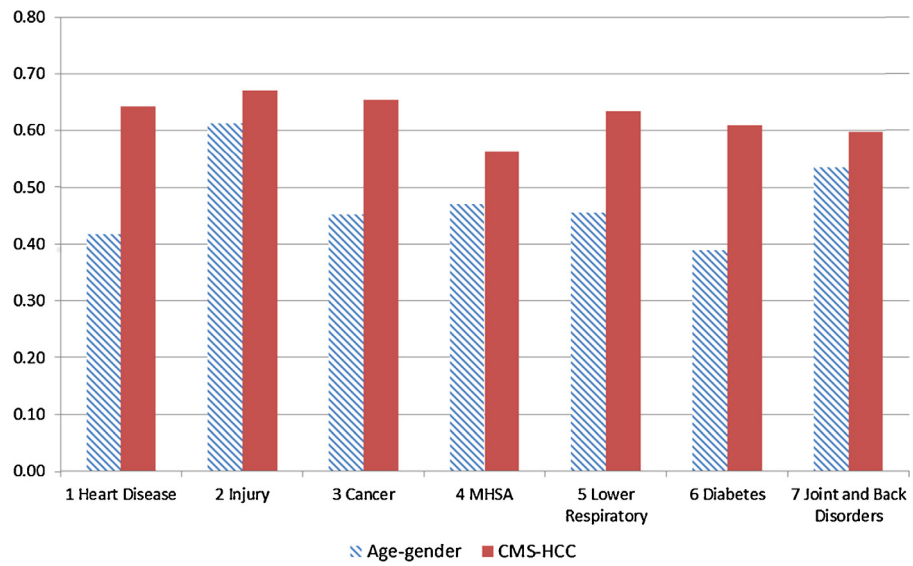

Fig. 2. Predictive ratio for CCS categories.

no systematic over or under-prediction for different parts of the distribution of spending. Results for the service-specific categories (not shown) also fell close to the 45-degree line.

6. Results: selection incentives

We first present results for selection incentives in terms of predictive ratios and the selection index with our basic empirical model. We then consider the robustness of our findings to alternative empirical approaches.

6.1. Predictive ratios

Fig. 2 shows the results of our analysis of service-level incentives using predictive ratios for each of the service categories. The predictive ratio for each of the two risk adjustment systems is the ratio of total revenue, premiums plus net risk adjustment payment from the Exchange, to total cost for all services for persons in each category.

For all seven groups the predictive ratio is considerably less than one in both risk adjustment systems. (For the population overall, the predictive ratio is 1.0, by definition, so there are obviously other population groups without any of these conditions with predictive ratios above one.) With age and sex adjustment, the predictive ratio exceeds 0.5 only for Injury, and Non-Traumatic Joint and Back Disorders. The CMS-HCC system moves all the predictive ratios toward 1.0. With HCCs, all groups get to 0.6 or above, with the exception of Mental Health and Substance Abuse. This is an indication that overall, the 20% of the patients with any one of the illness groups indicated in the figure are financially unattractive to the plan, and that the least attractive group includes enrollees with mental health and substance abuse.

6.2. Selection incentive index

The selection index I_s, from (2), is the product of demand elasticity of service s (ε_s), the coefficient of variation of predicted spending on service s (cv_s), and the correlation of service spending with gains and losses (ρ_s). Table 6 reports all three, and their product, for each of our service categories.

Elasticity of demand, as noted earlier, is not estimated in this paper. Instead we use estimates from the literature, shown in column (1) of Table 6. Our selection index is a relative measure, so only the relative values of demand elasticity across the services matter for our index. Only one of the services, mental health, has been subject to extensive separate study, with the general finding that the demand elasticity is roughly twice that for other forms of health care (Frank and McGuire, 2000). The demand elasticity of −0.2 for other services is based on the RAND Health Insurance Experiment (Newhouse, 1993), and we double this for our Mental Health and Substance Abuse Group.[53]

Column (2) reports our measure of predictability, the coefficient of variation (cv) of spending for each service. As noted earlier, this measure of predictability goes up as a population is better able to predict how their spending differs from the mean. The distribution of predicted spending comes from our two-part model of spending. Cancer, Mental Health, and Diabetes, all have coefficient of variations exceeding 5.0, indicating a high level of predictability of these services. Least predictable services include Injuries, and Joint and Back Disorders. These rankings accord with intuition. The advantage of a theory-driven metric is that not only the rankings matter, but the measure itself indicates the strength of this component of the incentives to select based on these services.

Column (3) reports predictiveness – the correlation between service spending in each category, and gains and losses based on the CMS-HCC risk adjustment model. Payments take into account the premium/risk adjustment system to be used in Exchanges. All correlations are negative, indicating that in all service categories, higher levels of spending on that category is correlated with losses. These correlations fall into two groups. Heart Disease, Injury, Cancer and Joint and Back Disorders have larger negative correlations

[53] Studies of demand response of mental health care to cost sharing in the post-managed care era find response about the same as other health care. With managed care, demand-side price is not always the binding constraint on use. The concept of "demand response" used in the selection index is not the empirical relationship between price and use as found in the presence of managed care rationing, but the "pure" demand response, the shape of the demand or perceived marginal benefit schedule, unadulterated by managed care. It is thus reasonable to use pre-managed care relative demand-response estimates in the index.

Table 6
Predictability and predictiveness of spending by CCS group.

	Demand elasticity	Predictability	Predictiveness	Selection index	
	(1)	(2)	(3)	(4)	(5)
1. Heart Disease	−0.2	2.97	−0.30	0.178	48
2. Injury	−0.2	1.52	−0.32	0.097	26
3. Cancer	−0.2	5.61	−0.33	0.370	100
4. Mental Health and Substance Abuse	−0.4	5.10	−0.16	0.326	88
5. Lower Respiratory Disorders	−0.2	4.27	−0.18	0.154	42
6. Diabetes	−0.2	5.92	−0.19	0.225	61
7. Non-Traumatic Joint and Back Disorders	−0.2	2.66	−0.30	0.160	43

Demand elasticity (1) is taken from the literature as explained in the text. Predictability (2) is the CV of predicted spending from the basic prediction model with age–sex, health and mental health status, and previous spending categories as regressors. Predictiveness (3) is the correlation of spending within a CCS group and gain/loss for the CMS-HCC risk adjustment system. Values for the selection index are −1 times the product of demand elasticity, predictability and predictiveness. The selection index is thus the product of the new (1), (2) and (3). Column (5) is the rescaled index.

(around −0.3), whereas Mental Health, Lower Respiratory Disorders and Diabetes have smaller ones. Because this measure is a correlation of spending with payments less costs at the enrollee level it reflects both the correlation of spending in one category with total spending, and how well spending is predicted by the risk adjustment/premium system. A service like Injury that is unpredictable will not be picked up well by a risk adjustment system, tending to increase the absolute value of the negative correlation.

Column (4) of Table 6 computes the value of I_s for each service, and column (5) simply rescales this for ease of comparison, setting the value at 100 for Cancer, the group with the highest value. Two of these services stand out: Cancer and Mental Health and Substance Abuse. Cancer stands out because it is predictable and predictive (of losses), and Mental Health stands out because it is predictable and has a higher demand elasticity (by assumption). If Mental Health and Substance Abuse were assumed to have the same demand elasticity as other services, the normalized index would fall back into the pack at 44 and the second-highest index service would be diabetes at 61.

6.3. Robustness

We checked the robustness of our predictability and predictiveness measures to three changes in assumptions. First, we modified the specification of our model of predicted spending in two ways. In the baseline model, the year 2 spending is predicted by applying the two-part quasi-likelihood generalized linear model (GLM) using the following year 1information: age, sex, self-rated health and mental health status, total spending, and CCS categorical spending. Instead of utilizing all the information available, the first alternative approach includes only age–sex combination as control variables. Although the absolute values of *cvs* differ from those in the baseline model, the relative scales remain largely unchanged. That is, the services with high *cvs* in the baseline model, e.g., cancer and diabetes, tend to have higher *cvs* than others in the alternative model. In the second alternative approach, we fit an OLS model instead of a GLM. The absolute values of *cvs* in both models take on the same relative values.

Second, we alter the risk adjustment model. In the baseline model, the CMS-HCC model is implemented to calculate risk adjustment weights, with age–sex combination and indicators of major medical conditions as control variables. The alternative approach includes only the age–sex combination. The predictiveness index remains almost the same in the alternative model.

Third, σ, the share of risk adjustment, is changed from 0.5 to 0.8 to see if it has an impact on the predictiveness index. Plan revenues are calculated based on the two shares and they are highly correlated with a correlation coefficient of 0.973, demonstrating that a change in the risk adjustment budget would have little effect on the predictiveness index. It of course has no effect at all on predictiveness or demand elasticity.

7. Discussion

Architects of the new Exchanges have taken steps to mitigate the problem of adverse selection, including requiring open enrollment, regulating the benefit package, risk adjusting plan payments, implementing risk corridors, and requiring a temporary reinsurance feature. This paper develops a method for assessing incentives for adverse selection that may remain even after these fixes. We make two primary contributions. First, we emphasize the role of premiums in plan revenues and incentives. This is critical in Exchanges where revenues per person can vary by a factor of three or greater. Taking account of premiums requires addressing how premiums will be determined in equilibrium. We make the conventional assumption here by assuming a competitive (zero-profit) equilibrium.[54] While natural, this assumption may not be correct. A limitation of our paper and a direction for future research would be to explore incentives in environments with imperfect competition among health plans (and possibly among providers).

Second, we show how to operationalize the implications of profit maximization for incentives to engage in service-level selection. Specifically, drawing on earlier papers showing that services that are both predictable and predictive are subject to underprovision, we measure both predictability and predictiveness and the consequent incentives to underprovide by major service area. This is an empirical task that requires simulating the basics of the payment system in Exchanges and data from a population similar to those who will likely be participating in an Exchange. These requirements call attention to two limitations of our analysis: first, we capture the major, but not all of the financial features of Exchange payment, and second, our data are from a national survey, not from actual Exchange participants.

[54] A constant markup due to market power is unlikely to affect the relative incentives for service level selection. A firm with market power may underprovide quality (here ration more tightly) as well as markup the price. Where market power is exercised depends on elasticities of demand, which may further differ by service. We suspect that the quality elasticity of demand will be lower than the price elasticity of demand, implying that market power will be exercised more in the form of lower quality. Pursuit of this idea is an important area for more theoretical and empirical investigation.

Profit incentives to plans are mitigated in the short term by reinsurance features and indefinitely by risk corridors that limit gains and losses. These features are likely to reduce but not eliminate service-level selection incentives (Zhu et al., forthcoming). Risk adjustment is done at the plan level (i.e., Bronze, Silver, etc.) in Exchanges, and our analysis assumed one plan level. Importantly, as we discussed above, data from the public-use files in MEPS do not incorporate the fineness of the risk adjustment systems. Also in terms of data, our sample size is low for risk adjustment modeling.

Another notable limitation is that we have assumed full compliance with the insurance mandate. Early enrollment in Exchanges did not go smoothly in late 2013, some states showed little enthusiasm for the policy, and nearly everywhere enrollment was slower than expected. Partial compliance may turn out to be uneven across population groups. If the mandate regulations are not very effective, groups with less to gain from participation, younger people and "better risks" generally may be less likely to participate. This will affect the overall size and vitality of the Exchanges as well as the mix of risks to be insured. A change in the composition of the risk pool due to noncompliance will make insurance more expensive on average, but it is unclear how it would affect incentives for selection for particular disease areas. Further, as noted in the introduction, most plans offered in the Exchanges will have narrower networks than current commercial plans. Although experience could well differ in such networks, it is hard to predict how, if at all, that would affect our conclusions.

With these qualifications in mind, we find, nonetheless, strong incentives to underprovide care to persons with some chronic illnesses may remain in spite of risk adjustment and other payment system features designed to mitigate against underprovision. We measure these incentives using an improved version of a predictive ratio and by a selection index derived from plan profit maximization. These measures can be readily applied to data, including data as it emerges from Exchange experience. While it is not surprising that plans have incentives to avoid sick people, our methods allow us to go beyond this general statement and identify the disease areas that should be of special concern.

Measured by predictive ratios, the strongest incentives are to discourage enrollment by people with mental health and substance abuse problems. Even though these disorders are themselves not very expensive, the people who use these services tend to use more of all other services, and in disease areas not tracked well by existing risk adjustment. By comparison, risk adjustment does relatively well in picking up the extra costs (across all diseases) for persons with cancer and diabetes.

Using the selection index based on profit maximization, however, takes into account the predictability of various illnesses – with more predictable conditions creating stronger incentives for a plan to use as a selection device. With this approach, cancer rises to the top in terms of disincentives to supply, with mental health and substance abuse as number two. Incentives to a plan to over or under-provide services depend on the patterns of disease in the underlying population, but these illnesses were also found to be subject to incentives to undersupply in Medicare (Ellis and McGuire, 2007).

Interestingly, cancer diagnoses were also among the least profitable diagnoses in the Medicare Advantage population studied by Newhouse et al. (2013). They examined 48 unique combinations of HCC's including single HCC's; all seven cancer diagnosis they examined were among the 12 lowest margin HCC's. The only mental health diagnosis they examined was major depressive, bipolar, or schizophrenia without another CMS-HCC category coded. Those individuals were around the median in profitability. The data suggested that both the ability to manage the disease medically and the market power of providers treating the disease mattered. These might differ for a Medicare Advantage population than for the populations in an Exchange. Importantly, Newhouse et al. found no evidence of selection despite substantial differences in margins across the categories. The distribution of the Medicare Advantage population across these HCC's was very close to that of the traditional Medicare population. Whether this is attributable to the effectiveness of Medicare regulations inhibiting selection or the costliness of selecting by disease or both is unknown.

As data begin to come in from the Exchanges, incentives for undersupply by disease area can be assessed more accurately. State health insurance regulators can be alert to underservice in disease areas, perhaps paying attention to the level of payment and the depth of the networks plans create for these conditions. A more drastic approach to an area subject to underservice is to regulate health insurance contracts, such as by "carving out" the benefit and writing a separate contract (perhaps at the state level for all Exchange participants) for supply of care in the designated disease area. Modification of the terms of the payment system, altering rules for risk adjustment or premium setting, or changing reinsurance rules, for example, will have differential effects on incentives in different disease areas. Incentive effects of these possible changes or other policy options can be examined with the methods developed here.

Acknowledgements

Research for this paper was supported by the National Institute of Mental Health (R01 MH094290) and the National Institute of Aging (P01 AG032952). This paper represents the views of the authors and no official endorsement by the Agency for Healthcare Research and Quality or the Department of Health and Human Services is intended or should be inferred.

Appendix. Equilibrium premiums depend on risk adjustment

This section describes the relationship between risk adjustment policy and premiums in an Exchange context. Let the total number of people be N and health care costs of individual i be x_i, with an overall average of $\bar{x}$. People vary in two observable dimensions, according to health status, the basis of risk adjustment, and according to another set of characteristics, the basis of premiums. Health status is indexed by h, $h = 1, \ldots, H$; premium characteristics are indexed by t, $t = 1, \ldots, T$. Each of these categorizations is mutually exclusive so that each person is characterized by an (h, t) pair. There can be overlap between the factors (e.g., age categories) used in classifying h and classifying t.

Define x_{ht} to be the average cost of person of type (h, t), and n_{ht} to be the number of people of type (h, t). Health care costs are plan costs (which must be covered by plan payments) and are fixed (do not depend on risk adjustment or premiums).

We further define:

$$n_h = \sum_t n_{ht} \quad \text{number of people with health status } h.$$

$$n_t = \sum_h n_{ht} \quad \text{number of people with personal characteristics } t.$$

$$x_h = \frac{1}{n_h} \sum_h n_{ht} x_{ht} \quad \text{average covered costs of person of health status } h.$$

$$x_t = \frac{1}{n_t} \sum_h n_{ht} x_{ht} \quad \text{average covered cost of person with personal characteristics } t.$$

All or some of the premiums paid to a plan will be subject to adjustment based on the health status (h) characteristics of the

Table A.1
Sample size and spending for HCCs included in the empirical analysis.

HCC	Obs	Mean	Std dev	Min	Max
1. HIV/AIDS	171	$3728	$7825	$0	$50,000
2. Septicemia/Shock	94	$3689	$4724	$0	$28,607
5. Opportunistic Infections	270	$3807	$6409	$0	$50,000
7. Metastatic Cancer and Acute Leukemia	25	$13,086	$18,868	$624	$50,000
9. Lymphatic, Head and Neck, Brain, and Other Major Cancers	21	$16,937	$20,721	$0	$50,000
10. Breast, Prostate, Colorectal and Other Cancers and Tumors	324	$5211	$8807	$0	$50,000
19. Diabetes without Complication	930	$6997	$10,436	$0	$50,000
27. Chronic Hepatitis	25	$6491	$9082	$0	$28,218
31. Intestinal Obstruction/Perforation	54	$5331	$10,308	$0	$50,000
32. Pancreatic Disease	23	$5052	$7662	$0	$28,417
33. Inflammatory Bowel Disease	35	$4824	$6459	$0	$25,189
37. Bone/Joint/Muscle Infections/Necrosis	139	$4918	$7961	$0	$50,000
38. Rheumatoid Arthritis and Inflammatory Connective Tissue Disease	162	$6664	$10,861	$0	$50,000
44. Severe Hematological Disorders	27	$6714	$13,379	$0	$50,000
52. Drug/Alcohol Dependence	27	$2712	$2694	$0	$11,130
55. Major Depressive, Bipolar, and Paranoid Disorders	88	$6802	$8628	$0	$50,000
72. Multiple Sclerosis	20	$16,866	$15,583	$0	$50,000
73. Parkinsons and Huntingtons Diseases	46	$6590	$6867	$0	$33,708
74. Seizure Disorders and Convulsions	36	$4033	$5531	$106	$22,236
75. Coma, Brain Compression/Anoxic Damage	805	$5879	$9490	$0	$50,000
77. Respirator Dependence/Tracheostomy Status	61	$2579	$3195	$0	$13,543
79. Cardio-Respiratory Failure and Shock	20	$6410	$12,586	$0	$50,000
80. Congestive Heart Failure	141	$8396	$11,769	$0	$50,000
82. Unstable Angina and Other Acute Ischemic Heart Disease	75	$9095	$12,970	$0	$50,000
83. Angina Pectoris/Old Myocardial Infarction	43	$9288	$11,932	$387	$50,000
92. Specified Heart Arrhythmias	98	$6971	$10,021	$0	$50,000
96. Ischemic or Unspecified Stroke	39	$9100	$13,777	$0	$50,000
104. Vascular Disease with Complications	201	$5913	$9262	$0	$50,000
105. Vascular Disease	32	$5443	$9345	$0	$50,000
108. Chronic Obstructive Pulmonary Disease	730	$4903	$8552	$0	$50,000
119. Proliferative Diabetic Retinopathy and Vitreous Hemorrhage	132	$5079	$8871	$0	$50,000
131. Renal Failure	21	$17,203	$18,894	$0	$50,000
148. Decubitus Ulcer of Skin	34	$5920	$8417	$0	$32,782
155. Major Head Injury	67	$3011	$7085	$0	$50,000
157. Vertebral Fractures without Spinal Cord Injury	226	$6517	$9561	$0	$50,000
158. Hip Fracture/Dislocation	20	$6470	$10,682	$0	$36,797
164. Major Complications of Medical Care and Trauma	44	$5126	$7002	$0	$34,621
176. Artificial Openings for Feeding or Elimination	939	$5820	$8810	$0	$50,000
High expense HCCs[a]	54	$12,344	$14,017	$0	$50,000
Low expense HCCs[b]	81	$5474	$7990	$0	$50,000
No HCC	16,233	$1351	$3728	$0	$50,000

Source: MEPS 2005–2009. All spending reported in $2009, $N = 20{,}865$.

[a] This category includes HCCs for which there are fewer than 20 observations in the category and mean spending is greater than $10,000 per year. It includes hcc8, hcc45, hcc107, hcc111 and hcc130.

[b] This category includes HCCs for which there are fewer than 20 observations in the category and mean spending is less than $10,000 per year. It includes hcc25, hcc26, hcc54, hcc68, hcc69, hcc70, hcc71, hcc95, hcc100, hcc132, hcc174 and hcc177.

persons who join. Suppose a share σ, $0 < \sigma \leq 1$, of costs were subject to risk adjustment. This would mean that the plan could be thought of paying in to the Exchange authority $\sigma\bar{x}$ for each enrollee, and getting back a risk adjusted payment dependent on h. Risk-adjusted payments sum to the σ share of costs:

$$\sum_h n_h r_h = \sigma N \bar{x} \tag{A.1}$$

Obviously, there are many risk adjustment systems, r_h, which satisfy this constraint (including one in which there is no risk adjustment at all and the plan receives a flat payment back for each person). We estimate the CMS-HCC model in the conventional way to find the relative weights for risk adjustment.

Assuming a plan draws a random distribution of enrollees, the zero-profit constraint for any premium type can thus be written:

$$p_t + \left\{ \frac{1}{n_t} \sum_h n_{ht} r_h - \sigma\bar{x} \right\} - x_t = 0 \quad \text{for each } t \tag{A.2}$$

The terms in brackets in (A.2) describe the risk adjustment. The plan sends in a share of average cost and gets back a risk-adjusted payment. This will be net positive or negative depending on whether the plan draws a sicker or healthier mix of enrollees. The plan pays average costs, x_t, for persons with premium type t. We use the series of equations in (A.2) to solve for the premiums p_t that just cover these costs, taking account of the risk adjustment in an Exchange.

We can see how (A.2) works in a couple of extreme cases. Suppose there is no Regulator and no risk adjustment. In this case, the terms in brackets in (A.2) are absent and $p_t = x_t$ for each premium type – this is the outcome of an unregulated competitive individual health insurance market. Alternatively, suppose $\sigma = 1$, and the Regulator risk adjusts all payments. Since risk adjustment categories are correlated with premium categories (age is in both, and HCCs are related to age), risk adjusting all costs will minimize the residual of costs to be picked up by premium categories. Premiums, p_t, will be "compressed" toward $\bar{x}$. The additional residual explanatory power of age and geography will increment the "fit" of the payment system over and above the R^2 of the risk adjustment

Table A.2
Age–sex and CMS-HCC risk adjustment model results (N = 20,865).

	Age–sex model		CMS-HCC model	
Variables	Parameter estimate	Std. error	Parameter estimate	Std. error
F0–5	$753	($267)	$446	($254)
F6–12	$1004	($249)	$798	($236)
F13–17	$1200	($208)	$927	($198)
F18–24	$1587	($137)	$1209	($131)
F25–34	$2356	($116)	$1901	($110)
F35–44	$2211	($113)	$1501	($108)
F45–54	$3198	($109)	$1960	($107)
F55–64	$4809	($140)	$2927	($142)
M0–5	$882	($268)	$506	($255)
M6–12	$756	($241)	$491	($229)
M13–17	$1108	($203)	$880	($192)
M18–24	$567	($125)	$387	($119)
M25–34	$853	($109)	$607	($104)
M35–44	$1399	($109)	$923	($104)
M45–54	$2706	($116)	$1744	($112)
M55–64	$4381	($147)	$2824	($144)
HCC1			$1498	($373)
HCC2			$718	($501)
HCC5			$751	($298)
HCC7			$7337	($974)
HCC9			$10,907	($1063)
HCC10			$1446	($274)
HCC19			$3379	($168)
HCC27			$2490	($969)
HCC31			$462	($661)
HCC32			$1319	($1010)
HCC33			$1693	($819)
HCC37			$1424	($413)
HCC38			$2791	($384)
HCC44			$797	($938)
HCC52			−$854	($935)
HCC55			$3436	($520)
HCC72			$13,491	($1083)
HCC73			$1928	($716)
HCC74			$1591	($806)
HCC75			$2270	($177)
HCC77			−$450	($621)
HCC79			$2380	($1083)
HCC80			$3052	($414)
HCC82			$3692	($565)
HCC83			$2992	($744)
HCC92			$1920	($493)
HCC96			$3572	($780)
HCC104			$1824	($345)
HCC105			−$126	($858)
HCC108			$1824	($184)
HCC119			$1436	($424)
HCC131			$9922	($1065)
HCC148			$1487	($832)
HCC155			$524	($593)
HCC157			$1908	($332)
HCC158			$3053	($1084)
HCC164			$912	($733)
HCC176			$2248	($165)
High-exp			$6235	($664)
Low-exp			$2156	($539)
R^2	0.186		0.272	

alone.[55] As we note in the text, fit of the payment system is not the only criterion that comes into play here. Premiums are prices to potential enrollees and influence sorting among plans. A maximally compressed premium schedule (aka "community rating") is unlikely to be best to encourage efficient sorting.

System (A.2) could be adapted to multiple plan types. Since the plan assessment is independent of plan premiums, this same formula could be applied to plans with more or less coverage, such as the bronze, silver, gold and platinum plans that will operate in Exchanges. The risk adjustment assessment could be on bronze plan costs (for example) and average costs in a plan would depend on plan type. Plans with more extensive coverage would need higher premiums to break even. The system (A.2) could also incorporate restrictions on rate bands as will apply in an Exchange. If, for example, young and old premiums are tied by a 1–3 ratio, we will require the pair of groups to break even and the premiums to be set in the required ratio. The presence of federal premium subsidies has no direct effect on the premiums necessary for the plans to break even (Tables A.1–A.3).

[55] This two-step fitting process is inferior to a one-step in which the risk adjustment weights are "chosen" simultaneously with market premiums. This point underlies the analysis in McGuire et al. (2013).

Table A.3
Market premiums under age–sex and CMS-HCC risk adjustment.

Premium category	*N*	Average cost	Average RA payment		Premium	
			Age–sex	CMS-HCC	Age–sex	CMS-HCC
Northeast, 0–18	375	$1138	$493	$564	$1694	$1623
Northeast, 19–34	839	$1404	$645	$685	$1809	$1770
Northeast, 35–54	1239	$2543	$1166	$1182	$2427	$2411
Northeast, 55–64	455	$4153	$2212	$2236	$2991	$2967
Midwest, 0–18	500	$1455	$498	$518	$2007	$1987
Midwest, 19–34	1245	$1718	$660	$709	$2108	$2059
Midwest, 35–54	1647	$2582	$1169	$1202	$2463	$2430
Midwest, 55–64	585	$4859	$2230	$2303	$3679	$3606
South, 0–18	1105	$728	$489	$475	$1289	$1303
South, 19–34	2598	$1245	$674	$666	$1621	$1629
South, 35–54	3277	$2190	$1156	$1155	$2085	$2085
South, 55–64	1086	$4774	$2209	$2317	$3615	$3507
West, 0–18	838	$926	$489	$468	$1486	$1507
West, 19–34	2047	$1169	$647	$602	$1572	$1617
West, 35–54	2324	$2093	$1145	$1083	$1998	$2060
West, 55–64	705	$3935	$2207	$2089	$2778	$2896

Calculations assume $\sigma = 0.5$ so 0.5 of average cost ($1050) is paid into an Exchange Authority for each enrollee. This is risk adjusted by one of two systems and funds returned to plans.

References

Agency for Healthcare Research and Quality, 2008. MEPS HC-092 1996–2004 Risk Adjustment Scores, Public Use File. Gaithersburg, MD.

Agency for Healthcare Research and Quality, 2011. MEPS HC-128: Medical Conditions File. Rockville, MD.

Barry, C.L., Weiner, J.P., Lemke, K., Busch, S.H., 2012. Risk adjustment in health insurance exchanges for individuals with mental illness. American Journal of Psychiatry 169, 704–709.

Barry, C.L., Ridgely, M.S., 2008. Mental health and substance abuse insurance parity for federal employees: how did health plans respond? Journal of Policy Analysis and Management 27 (1), 155–170.

Blumberg, L., Pollitz, K., 2009. Health Insurance Exchanges: Organizing Health Insurance Marketplaces to Promote Health Reform Goals. The Urban Institute, Washington, DC.

Breyer, F., Bundorf, M.K., Pauly, M.V., 2012. Health care spending risk, health insurance, and payment to health plans. In: Pauly, McGuire, Barros (Eds.), Handbook of Health Economics, vol. 2. Elsevier, Amsterdam, pp. 691–762.

Cao, Z., McGuire, T.G., 2003. Service-level selection by HMOs in Medicare. Journal of Health Economics 22 (6), 915–931.

Chandra, A., Gruber, J., McKnight, R., 2011. The importance of the individual mandate – evidence from Massachusetts. New England Journal of Medicine 364 (4), 293–295.

Collins, S.R., Garber, T., 2013. The Affordable Care Act's Health Insurance Marketplaces: A Progress Report. The Commonwealth Fund Blog, February 21, Available online at: http://www.commonwealthfund.org/Blog/2011/Jun/State-Health-Insurance-Exchange-Legislation.aspx (accessed 27.02.13).

Congressional Budget Office, 2011. CBO's Analysis of the Major Health Care Legislation Enacted in March 2010. Testimony by Douglas W. Elmendorf before the Subcommittee on Health, Committee on Energy and Commerce, US House of Representatives, March 30, 2011.

Cook, B.L., McGuire, T.G., Meara, E., Zaslavsky, A.M., 2009a. Adjusting for health status in non-linear models of health care disparities. Health Services Outcomes Research and Methodology 9 (1), 1–21.

Cook, B.L., McGuire, T.G., Zuvekas, S.H., 2009b. Measuring trends in racial/ethnic health care disparities. Medical Care Research and Review 66 (1), 23–48.

Cutler, D., Lincoln, G., Zeckhauser, R., 2010. Selection stories: understanding movement across health plans. Journal of Health Economics 29 (6), 821–838.

Cutler, D., Reber, S., 1998. Paying for health insurance: the tradeoff between competition and adverse selection. Quarterly Journal of Economics 113 (2), 433–466.

Cutler, D., Zeckhauser, R., 2000. The anatomy of health insurance. In: Culyer, A., Newhouse, J. (Eds.), Handbook of Health Economics, vol. I. Elsevier, Amsterdam.

Department of Health and Human Services, 2012. Patient Protection and Affordable Care Act; HHS notice of benefit and payment parameters for 2014; proposed rule. Federal Register 77 (236), 73117–73218.

Department of Health and Human Services, 2013. Patient Protection and Affordable Care Act, HHS notice of benefit and payment parameters for 2014 and amendments to the HHS notice of benefit and payment parameters for 2014; final rules. Federal Register 78 (47), 15410–15541.

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

Ellis, R.P., McGuire, T.G., 2007. predictability and predictiveness in health care spending. Journal of Health Economics 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? Applied Economics 45 (13), 1701–1712.

Einav, L., Finkelstein, A., 2011. Selection in insurance markets: theory and empirics in pictures. Journal of Economic Perspectives 25 (1), 115–138.

Ericson, K., Starc, A., 2012. Pricing Regulation and Imperfect Competition on the Massachusetts Health Insurance Exchange. In: NBER Working Paper #18089, Unpublished.

Foote, S.M., Jones, S.B., 1999. Consumer-choice markets: lessons from FEHBP mental health coverage. Health Affairs (Millwood) 18 (5), 125–130.

Frank, R.G., McGuire, T.G., 2000. Economics and mental health. In: Culyer, Newhouse (Eds.), The Handbook of Health Economics, vol. 1. Elsevier, Amsterdam.

Fullerton, C., Epstein, A., Frank, R., Normand, S.-L., Fu, C., McGuire, T., 2012. Medication use and spending trends among children with ADHD in Florida's Medicaid Program, 1996–2005. Psychiatric Services 63 (2), 115–121.

General Accounting Office, 2011. Medicare Advantage: Changes Improved Accuracy of Risk Adjustment for Certain Beneficiaries. GAO-12-52. December.

Glazer, J., McGuire, T., 2000. Optimal risk adjustment of health insurance premiums: an application to managed care. American Economic Review 90 (4), 1055–1071.

Hill, S., Zuvekas, S., Zodet, M., 2012. Validity of reported medicare Part D enrollment in the medical expenditure panel survey. Medical Care Research and Review 69 (6), 537–550.

Machlin, S., Cohen, J., Elixhauser, A., Beauregard, K., Steiner, C., 2009. Sensitivity of household reported medical conditions in the medical expenditure panel survey. Medical Care 47, 618–625.

McGuire, T.G., Glazer, J., Newhouse, J.P., Normand, S.-L., Shi, J., Sinaiko, A.D., Zuvekas, S., 2013. Integrating risk adjustment and enrollee premiums in health plan payment. Journal of Health Economics 32 (6), 1263–1277.

Newhouse, J.P., 1993. Free-For-All: Health Insurance, Medical Costs, and Health Outcomes: The Results of the Health Insurance Experiment. Harvard University Press, Cambridge, MA.

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 Financing Review 10 (3), 41–54.

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? Journal of Health Economics 32 (6), 1278–1288.

Padgett, D.K., Patrick, C., Burns, B.J., Schlesinger, H.J., Cohen, J., 1993. The effect of insurance benefit changes on use of child and adolescent outpatient mental health services. Medical Care 31 (2), 96–110.

Pauly, M., 1985. What is adverse about adverse selection? In: Scheffler, R.M., Rossiter, L.F. (Eds.), Advances in Health Economics and Health Services Research: Biased Selection in Health Care Markets. JAI Press, Greenwich, CT.

Pear, R., 2013. Lower health insurance premiums to come at cost of fewer choices. New York Times, page 1, September 22.

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

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

Trish, E., Damico, A., Claxton, G., Levitt, L., Garfield, R., 2011. A Profile of Health Insurance Exchange Enrollees. Kaiser Family Foundation. March.

Weiner, J.P., Trish, E., Abrams, C., Lemke, K., 2012. Adjusting for risk selection in state insurance exchanges will be critically important and feasible, but not easy. Health Affairs 31 (2), 306–315.
Wicks, E.K., Hall, M.A., 2000. Purchasing cooperatives for small employers: performance and prospects. Milbank Quarterly 78 (4), 511–546, iii.
Winkelman, R., Pepper, J., Holland, P., Mehmud, S., Woolman, J., 2011. Analysis of HHS proposed rules on reinsurance, risk corridors and risk adjustment. State Health Reform Assistance Network.
Zhu, J., Layton, T., Sinaiko, A., McGuire, T., 2014. The power of reinsurance in health insurance exchanges to improve fit of the payment system and reduce incentives for adverse selection. Inquiry (forthcoming).
Zuvekas, S.H., Olin, G., 2009. An examination of the accuracy of medicare expenditures in the medical expenditure panel survey. Inquiry 46 (1), 92–108.