Journal of Public Economics 84 (2002) 153–173
JOURNAL OF PUBLIC ECONOMICS
www.elsevier.com/locate/econbase

Setting health plan premiums to ensure efficient quality in health care: minimum variance optimal risk adjustment

Jacob Glazer[a,*], Thomas G. McGuire[b]

[a]*Tel Aviv University, Faculty of Management, Tel Aviv, Israel*
[b]*Harvard Medical School, Boston, USA*

Abstract

Risk adjustment refers to the practice of paying health plans a premium per person (or per family) based on a formula using risk adjusters, such as age or gender, and weights on those adjusters. One role of risk adjustment is to make sure plans have an incentive to accept all potential enrollees. Another role, at least as important in our view, is to lead health plans to choose the efficient level of quality of care for the various services they offer. Most of the research and policy literature on risk adjustment focuses on the first problem. This paper proposes a new way to calculate weights in a risk adjustment formula that contends with both problems. For a given set of adjusters, we identify the weights that minimize the variance in plan predictable health care costs that are not explained by risk adjustment (addressing the access problem), subject to the payments satisfying conditions for an optimal risk adjuster (making sure plans provide the efficient quality). We call the formula minimum variance optimal risk adjustment (MVORA).

1. Introduction

Managed health care plans – insurance and service organizations responsible for providing ‘medically necessary’ health care – enroll about 75 percent of the U.S. population and large shares of European populations. These health plans receive

*Corresponding author.
E-mail address: glazer@post.tau.ac.il (J. Glazer).

0047-2727/02/$ – see front matter
PII: S0047-2727(01)00123-2

individual or family premiums, paid by governments (e.g. Medicare or Medicaid in the U.S.), or by a combination of employers and consumers in U.S. employment-based health insurance. One of the major concerns with the health insurance/managed care health care policy is adverse selection (Enthoven, 1993). Generally, plans may take actions to discourage or encourage potential enrollees. For one thing, they may refuse some applicants, although overt actions to discourage individuals are normally prohibited and may be readily monitored. More troublesome is that plans may distort the mix of the quality of health care they offer to discourage high-cost persons from joining the plan. As a number of papers have observed, decisions about what care is medically necessary are fundamentally outside the scope of direct regulation (Miller and Luft, 1997; Newhouse, 1996). In this paper, we consider how risk adjustment of premiums paid to health plans can address what we refer to as the individual access problem and the quality problem.[1]

Risk adjustment refers to the practice of paying health plans a premium for each person or family based on a formula using risk adjuster variables, such as age, gender, or indicators of prior health care use, and weights on those adjusters. Risk adjustment researchers have sought to find the formula that does the best job of 'fitting' the distribution of health care costs in a population. Risk adjuster variables are chosen that are likely to be related to costs, and weights on those variables are chosen by regression techniques to minimize the sum of the squared residuals, a practice we will refer to here as 'conventional risk adjustment.'[2] The intuitive appeal of regression coefficients as weights on risk adjusters derives from a desire to address the individual access problem. Matching payments with individual costs as closely as possible may discourage plans from denying membership to some and aggressively recruiting others.

When the problem being addressed is plans' distortion of the quality of services

[1] A comprehensive discussion of health insurance premiums and risk adjustment would include analysis of the setting of the premiums charged to enrollees. Except in some individual insurance markets, individuals receive some 'subsidy' of this premium. In the case of government programs in the U.S. and some European countries, this subsidy is complete and the person or family pays nothing. In the employment-related health insurance context in the U.S., the employer pays a portion of the premium. Authors have considered how to set the consumer contribution in order to give consumers the incentives to join the right plans. This issue can also be considered within an adverse selection framework and may involve risk adjusting consumer premium contributions. The most immediate effect of the person-paid component of premiums is to influence the consumer's choice of plan. In this paper, as will be made clear in the next section, we will analyze a situation in which the consumer's contribution is zero. This is empirically accurate for many people in the U.S., including a large share of the employed population, as well as the entire population enrolling in managed health care in Medicare and Medicaid. It is also common in Europe.

[2] Age and gender are commonly used in regression formulae, but alone explain only about 1% of the variance in actual costs. Research on improving the explanatory power of these regressions has focussed on using sophisticated clinical algorithms to define new variables based on the diagnoses from prior health care use. Researchers have been able to run regressions explaining up to 10% of the variance in actual costs, but risk adjustment systems presently in use explain much less. See van de Ven and Ellis (2000) for a review of the empirical literature. Keenan et al. (2001) describes risk adjustment practices by major public and private payers in the U.S.

to affect the decisions of groups of potential applicants, however, the regression-based approach lacks intuitive foundation. Figuring risk adjustment weights to mitigate quality distortions requires a conception of how health plans set the quality of services, and a conception of what 'optimal' quality means in this context. In this paper we propose an approach to deriving weights on risk adjuster variables that begins with an explicit statement of health plan behavior in a market with heterogeneity in the demand for health care services, and an explicit statement of the economic efficiency problem.

With two efficiency problems in view, individual access and quality distortions, and one policy instrument, the formula for the premium paid to plans, it is natural to expect there to be some trade-off in meeting the two competing objectives. Although it has inspired most empirical risk adjustment research, 'individual access' to health plans is not the major social efficiency problem in the health plan market. In the U.S., Europe, Israel and Latin America, governments and employers require contracting plans to offer periodic 'open enrollments'. During open enrollment periods (once a year for most U.S. employers, and every month in Medicare), plans must accept any applicant. It seems clear that open enrollment regulation works well to ensure access. Simulation research demonstrates that existing risk adjustment systems leave some individuals as big 'losers' or winners (Chapman, 1997; Shen and Ellis, 2000), but there is no evidence that plans act to deny membership to individuals. By way of contrast, quality of care in managed care health plans is the major policy focus in the U.S. and elsewhere. We believe that the quality distortion problem is more important and less amenable to other regulatory solutions than the individual access problem, and is therefore the proper primary target of risk adjustment policy. Therefore, in the analysis below, when it comes to a tradeoff between policy objectives, we put more weight on maintaining quality than on individual access.

Plans' incentives to overprovide the quality of some services and underprovide the quality of others derives from the distribution of health care demands in the population. With data about the distribution of those demands, the regulator can anticipate plans' incentives, and impose a system of corrective taxes and subsidies at the person level, using the many observable variables available for each individual. Within this framework we derive a simple rule that characterizes optimal risk adjustment: the relation between the covariance between spending on a service and the risk adjusted premium and the covariance between spending on that service and the sum of spending on all services must be the same for all services that the plan provides. In general, there will be many combinations of weights on risk adjuster variables that satisfy the rule for optimal risk adjustment. By choosing from among the set of optimal risk adjustment weights the combination that leads to the best fit of premiums to costs, we integrate our approach with traditional regression-based methods, in effect designing a risk adjustment formula that contends with the two forms of adverse selection problems identified above. We will refer to our proposed method for risk adjustment as minimum variance optimal risk adjustment (MVORA), 'optimal' because it solves the quality distortion problem, and 'minimum variance' because it does so in a way that minimizes the sum of the squares of the deviations between premiums and costs at the individual level. In line with the literature on risk adjustment, we argue that this statistical criterion is motivated by a concern for individual access.

Our paper is related to several lines of research in health and public economics. Many papers in health economics are concerned with efficiency problems due to selection in insurance markets (see Cutler and Zeckhauser, 2000; Encinosa, 1999 and van de Ven and Ellis, 2000 for reviews). Researchers have long been aware that one significant adverse selection-related problem associated with competition among health plans is service competition to attract the good risks/deter the bad (Newhouse, 1996). Little research has been done, however, on the implications of this concern for a risk adjustment formula. So far, risk adjustment of this type has only been characterized in simple cases. Glazer and McGuire (2000) consider the two-service case and one risk adjuster, and show that the best risk adjustment weights are the solution to a pair of equations. The weights obtained by this procedure are generally different from those obtained from regression coefficients.

From the perspective of public economics, it is natural to view risk adjustment as question of optimal taxes and subsidies on the prices paid for health insurance, a framework adopted in several papers (Glazer and McGuire, 2000; Neudeck and Podczeck, 1996; Selden, 1999). By explicit characterization of the distortions emerging in markets for health plans, we are able to identify, Pigouvian fashion, the set of taxes and subsidies necessary to align private incentives to maximize profit with the social objective of production of the efficient quality of health care.

In the context of risk adjustment, the taxes and subsidies to correct incentives work in a unusual fashion. Normally, a tax or subsidy applies directly to the activity intended to be affected: for example, a tax on the volume of pollution is intended to reduce pollution. Obviously though, there are other ways to tax/subsidize to hit a particular pollution target. If the actual level of pollution were not verifiable, for example, a regulator familiar with the effect of pollution (on, say water quality) could put a tax on this effect in order to manipulate the firm into choosing the target level of pollution. The idea in this paper is the same. Quality provided by health plans is usually not verifiable and, hence, cannot be taxed or subsidized directly. However, the regulator can magnify or diminish the revenue consequences to the health plan's quality choices by choice of the risk adjustment formula. If older people value and join a plan in response to good care for cardiac problems and the regulator is concerned that the quality offered for these diseases is too low, the regulator can induce higher quality by 'subsidizing' older people.[3]

[3]The health payment literature has used this insight previously, as in Rogerson (1994) where the optimal hospital per discharge payment was figured in order to induce the desired non-verifiable quality. In this case, quality is one-dimensional, and the level of the payment was the only instrument necessary. Another early paper along the same lines is Ma (1994).

A central insight of this paper is that there is an instruments/targets feature of risk adjustment policy. A plan sets quality for many services. The regulator has many variables in a risk adjustment formula. When the regulator has more instruments than it needs, the quality problem can be addressed. The final choice of risk adjustment formula can then be made in order to contend also with individual access.

The paper is organized as follows: Section 2 presents our basic model of health plan behavior in an environment with no uncertainty and symmetric information. In Section 3 we present the conditions for optimal risk adjustment and in Section 4 we solve for the Minimum Variance Optimal Risk Adjustment formula. An example of how to figure MVORA is contained in Section 5. Section 6 extends our analysis to the case of uncertainty and asymmetric information.

2. Health plan behavior

Assume there are N individuals. Each one of them is about to choose a health plan. In this section, we will analyze the behavior of one (representative) health plan, taking the behavior of the others as given. In Section 3 we will analyze this behavior within a symmetric equilibrium. The model presented in this section is based on that in Frank et al. (2000), hereafter referred to as FGM. The health plan is paid a premium (possibly risk-adjusted) for each individual that enrolls. Individuals differ in their need/demand for health care, and choose a plan which maximizes their expected utility. 'Health care' is not a single commodity, but a set of services – maternity, mental health, emergency care, cardiac care, and so on. A health plan chooses a rationing or allocation rule for each service. The plan's choice of rules will affect which individuals find the plan attractive and will therefore determine the plan's revenue and costs. We assume that the plan must accept every applicant, and we are interested in characterizing the plan's incentives to ration services.

2.1. Utility and plan choice

The health plan offers S services. Let m_{is} denote the amount the plan will spend on providing service s to individual i, if he joins the plan, and let: $m_i = \{m_{i1}, m_{i2}, \ldots, m_{iS}\}$. The dollar value of the benefits individual i gets from a plan, $u_i(m_i)$, is composed of two parts, a valuation of the services an individual gets from the plan, and a component of valuation that is independent of services. We assume these enter additively in utility. Thus,

$$u_i(m_i) = v_i(m_i) + \mu_i \tag{1}$$

where,

$$v_i(m_i) = \sum_s v_{is}(m_{is})$$

is the service-related part of the valuation and is itself composed of the sum of the individual's valuations of all services offered by the plan. $v_{is}(\cdot)$ is the individual's valuation of spending on service s, also measured in dollars, where $v'_{is} > 0$, $v''_{is} < 0$. Assume for the moment that the individual knows $v_i(m_i)$ with certainty. (This assumption will be relaxed in Section 6). The non-service component is μ_i, an individual-specific factor (e.g. distance or convenience) affecting individual i's valuation, known to person i. From the point of view of the plan, μ_i is unknown, but is drawn from a distribution $\Phi_i(\mu_i)$. We assume that the premium the plan receives has been predetermined and is not part of the strategy the plan uses to influence selection.

The plan will be chosen by individual i if $u_i > \bar{u}_i$, where $\bar{u}_i$ is the valuation the individual places on the next preferred plan. We analyze the behavior of a plan which regards the behavior of all other plans as given, so that $\bar{u}_i$ can be regarded as fixed. Given m_i and $\bar{u}_i$, individual i chooses the plan if:

$$\mu_i > \bar{u}_i - v_i(m_i).$$

For now, we assume that, for each i, the plan has exactly the same information as individual i regarding the individual's service-related valuation of its services, v_i, and regarding the utility from the next preferred plan, $\bar{u}_i$. For each individual i, the plan does not know the true value of μ_i but it knows the distribution from which it is drawn. Therefore, for a given m_i and $\bar{u}_i$, the probability that individual i chooses the plan, from the point of view of the plan, is:[4]

$$n_i(m_i) = 1 - \Phi_i(\bar{u}_i - v_i(m_i)) \tag{2}$$

2.2. Managed care

Managed care rations the amount of health care a patient receives. Following Keeler et al. (1998) and FGM, let q_s be the service-specific shadow price the plan sets determining access to care for service s. A patient with a benefit function for service s of $v_{is}(\cdot)$ will receive a quantity of services, m_{is} determined by:

$$v'_{is}(m_{is}) = q_s \tag{3}$$

Let the amount of spending determined by the equation above be denoted by $m_{is}(q_s)$. Note that Eq. (3) is simply a demand function, relating the quantity of services to the (shadow) price in a managed care plan.

[4] An alternative interpretation is that index i describes a group of people with the same $v_i(m_i)$ function and $n_i(m_i)$ is then the share of this group that joins the plan.

2.3. Profit and profit maximization

Let $q=\{q_1,q_2,\ldots,q_S\}$ be a vector of shadow prices the plan chooses and $m_i(q)=\{m_{i1}(q_2),m_{i2}(q_2),\ldots,m_{iS}(q_S)\}$ be the vector of spending individual i gets by joining the plan. Define $n_i(q)\equiv n_i(m_i(q))$. Expected profit, $\pi(q)$, to the plan will depend on the individuals the plan expects to be members, the revenue the plan gets for enrolling these people, and the costs of each member.

$$\pi(q)=\sum_i n_i(q)[r_i-\sum_s m_{is}(q_s)] \tag{4}$$

where r_i is the risk-adjusted revenue the plan receives for individual i. The plan will choose a vector of shadow prices to maximize expected profit, Eq. (4). Defining $M_i=\Sigma_s\ m_{is}(q_s)$ to be total spending on a person, profit per person is $\pi_i=r_i-M_i$. Assuming that individuals share the same elasticity of demand for any service (allowing common elasticities to differ across services)[5], the profit maximizing condition for q_s, $s=1,2,\ldots,S$, becomes (see FGM):

$$q_s=\frac{\sum_i n_i m_{is}}{\sum_i \Phi_i' m_{is}\pi_i}. \tag{5}$$

2.4. The efficiency criterion

The health plan allocates resources efficiently when it acts so as to equalize the degree of rationing across all S service areas. This is (second-best) efficient in the sense that the value of a dollar of health care spending is equalized across all possible uses. Formally, this requires $q_s=q$, for all s, where q is some constant characterizing the overall degree of rationing. When $q=1$, a first-best efficiency is achieved. When $q>1$, second-best efficiency is achieved but the overall budget for health is 'too small,' and when $q<1$, it is 'too big.'

There is, however, no particular reason to expect profit maximization represented by Eq. (5) to lead to the same shadow price for each service s, unless the risk adjustment system is able to equalize the relative incentives to supply each service.

3. Optimal risk adjustment

In the analysis of plan behavior just discussed, the risk adjustment formula was taken as given. We now consider how to structure the risk adjustment formula to

[5]Note that the fact that all individuals share the same elasticity of demand for a certain service does not imply that their demand curves are identical.

induce the plan to provide services at the (second-best) efficient quality. A risk adjustment formula that achieves this goal will be referred to as *optimal risk adjustment*.

The analysis carried out here is very much in the spirit of the standard principal-agent literature. In that literature, the principal constructs an incentive scheme so that the agent, acting in its own interest, will behave according to the principal's goals. In what follows the regulator (the principal) sets a risk adjustment formula (an incentive scheme) to induce the profit maximizing plan (the agent) to provide the efficient quality (the regulator's goal).

We have in mind a situation where there are several identical health plans. Thus, Eq. (5) above describes the profit maximizing behavior of each of the health plans in equilibrium (taking the behavior of the other plans as given).

In a symmetric equilibrium, each individual has the same probability of being in the plan, so $n_i = n$. Assume that Φ_i is uniform and the same for all i. Then, we can simplify Eq. (5) to say that second-best efficiency requires that for each s, Eq. (6) holds.

$$\frac{\sum_i m_{is}}{\sum_i m_{is} r_i - \sum_i m_{is} M_i} = \lambda, \tag{6}$$

for some $\lambda > 0$.

Eq. (6) can be rewritten as

$$\frac{\operatorname{cov}(m_s,r) - \operatorname{cov}(m_s,M)}{\bar{m}_s} = (\bar{M} - \bar{r} + \frac{1}{\lambda}) \tag{7}$$

where covariances are defined as

$$\begin{aligned} \operatorname{cov}(m_s,r) &= \frac{1}{N}\sum_i (m_{is} - \bar{m}_s)(r_i - \bar{r}) \quad \text{and} \\ \operatorname{cov}(m_s,M) &= \frac{1}{N}\sum_i (m_{is} - \bar{m}_s)(M_i - \bar{M}), \end{aligned}$$

with

$$\frac{1}{N}\sum_i m_{is} \equiv \bar{m}_s, \bar{M} \equiv \frac{1}{N}\sum_i M_i, \text{and } \bar{r} \equiv \frac{1}{N}\sum_i r_i.$$

Eq. (7) embodies the main result of this paper. To equalize incentives to ration all services, the covariance of the risk adjuster with the use of every service must track the covariance of total predicted costs with the service's use. Intuitively, the optimal risk adjustment formula must have the property that by spending on a service, the cost consequences to a plan (from serving the demands of new members) relates to the revenue consequences (from the premiums they bring in) in the same way across all services. When this set of conditions holds, the plan has an incentive to ration all services with the same stringency.

It is worthwhile to contrast the conditions represented by Eq. (7) for the optimal way to set premiums to the usual approach based on statistical fit. Conventional risk adjustment seeks to bring per person premiums as close as possible to per person costs. The first contrast to be made is that we have S conditions characterizing optimal risk adjustment rather than one criterion based on fit. If a plan chooses the quality of S services, it is necessary to address the conditions for efficiency for each service, so more than one criterion seems evidently necessary. Second, the statistical criterion of fit of premiums, r, to costs, M, is replaced by criteria represented by covariances that lead profit maximization decisions of health plans to be identical with conditions for efficiency. When the extra revenue brought in by spending on service s (captured by the covariance between m_s and r) is in the same relation to the extra total costs incurred by spending on service s (captured by the covariance between m_s and M) for all services, profit-maximization is efficient.

We are now ready to consider how to use a risk adjustment formula to set premiums r in order to satisfy conditions for optimal risk adjustment. Suppose that a set of J risk adjusters (variables) is available to use in a risk adjustment formula. Call this set $X = (X_1, \ldots, X_J)$. The variable x_{ij} represents the value of the adjuster j for person i (x_{ij} could be the age or gender or health status of individual i). r_i is the risk-adjusted premium payment made on behalf of person i. A risk adjusted premium can be written as

$$r_i = X_i \beta \tag{8}$$

where β is a vector of weights on the risk adjuster variables. As can be seen from Eq. (8), for a given set of risk adjuster variables, the issue of designing a risk adjustment formula is a matter of setting the right weights β. Rather than using a regression, we find the β's as solutions to optimality conditions.

Using Eq. (8), each of the S equations in Eq. (7) can be written as

$$\frac{\sum_j \beta_j \text{cov}(m_s, x_j) - \text{cov}(m_s, M)}{\bar{m}_s} = \bar{M} - \bar{r} + \frac{1}{\lambda} \tag{9}$$

Note that Eq. (9) is linear in the coefficients of the risk adjusters β_j. The RHS of Eq. (9) is the same for all S equations.

It is interesting to note what Eq. (9) implies for the relationship between the covariance of the x's with expected costs on a service and the β's. If the $cov(m_s,x)$ is small, in order to equalize this to a target covariance, the β's must be larger.[6]

Solving for $\bar{M} - r + 1/\lambda$, we are left with $S-1$ linear restrictions on the β_j's. Each of these restrictions takes the form of Eq. (10). In addition, a given budget b in the form of an average risk adjusted premium yields another linear restriction. We define *optimal risk adjustment* to be the weights $\beta_1, \ldots, \beta_J$ that satisfy the set of $S-1$ equations (10), and the budget constraint in Eq. (11). We label any set of β's satisfying these s conditions $\beta_j^{\,o}$.

$$\frac{\sum_j \beta_j^{\,o} \operatorname{cov}(m_s,x_j) - \operatorname{cov}(m_s,M)}{\bar{m}_s} = \frac{\sum_j \beta_j^{\,o} \operatorname{cov}(m_S,x_j) - \operatorname{cov}(m_S,M)}{\bar{m}_S} \qquad (10)$$

$$s = 1, \ldots, S-1$$

$$\frac{1}{N}\sum_j \sum_i \beta_j^{\,o} x_{ij} \equiv \sum_j \beta_j^{\,o} \bar{x}_j = b \qquad (11)$$

The system of Eqs. (10) and (11) will have at least one solution if a rank condition is satisfied, here, if $J > S$. We have assumed that the plan makes S independent rationing decisions, and there are J risk adjusters, none of which is a linear combination of the others. It is unclear what should be said about the magnitude of S, the number of rationing decisions made by a plan. On the one hand, a plan offers a very large number of different services. But on the other hand, from the standpoint of what management may practically pay attention to, there may be many fewer decision variables. For purposes of management, services might be grouped in just a few categories: rationing may be done for example at the level of primary care, pediatrics, mental health, cardiac care, and so on, leading to perhaps 10–20 groups. (There are 19 Major Diagnostic Categories, MDCs, for instance, that underlie the DRG system). As for the number of risk adjusters, J, this number could be very large. Consider age and gender alone. While only two 'variables' there are many potential degrees of freedom in age and gender. Medicare exploits these by creating age-gender cells. In a population ranged in age from say 1–65, a payer might feasibly create ten age $\times$ two gender cells for 20 mutually exclusive variables to use for risk adjustment. Adding zone of residence and eligibility status (among working people: employee, spouse, other dependent), geometrically enlarges J. Furthermore, diagnoses from prior health care use have been grouped into scores of categories in recently developed diagnosis-based classification systems (Ellis et al., 1996; Weiner et al., 1996). It

[6] This is a generalization of a finding of Glazer and McGuire (2000), where it was shown that with one signal (risk adjuster variable), as the informativeness of the signal deteriorates, the weight on the signal increases in the optimal risk adjustment formula.

thus seems quite safe to proceed on the assumption that the number of risk adjuster variables, J, exceeds the number of decisions made by management, S.

4. Minimum variance optimal risk adjustment (MVORA)

For a given per person budget b and a set X of risk adjusters, *Minimum Variance Optimal Risk Adjustment* (MVORA) is a vector of risk adjustment weights $\beta^* = (\beta_1^*, \ldots, \beta_J^*)$ that solves the following (constrained minimization) problem:

$$\text{Minimize}_{\beta_1, \ldots, \beta_J} \ \frac{1}{N} \sum_i \left(M_i - \sum_j \beta_j x_{ij}\right)^2 \tag{12}$$

$$\text{s.t.} \quad \sum_j \beta_j \left[\frac{\text{cov}(m_s, x_j)}{\bar{m}_s} - \frac{\text{cov}(m_S, x_j)}{\bar{m}_S} \right] = \frac{\text{cov}(m_s, M)}{\bar{m}_s} - \frac{\text{cov}(m_S, M)}{\bar{m}_S}$$

$$s = 1, \ldots, S-1 \tag{13}$$

$$\text{and} \quad \sum_j \beta_j \bar{x}_j = b \tag{14}$$

The set of $S-1$ equations in Eq. (13) is a rewriting of the $S-1$ equations in Eq. (10).

MVORA contends with both problems caused by adverse selection. The constraints guarantee the risk adjustment formula is optimal, in the sense that no quality distortion will take place. Minimizing the sum of the square of the deviations addresses problems that may arise when payments deviate from costs at the individual level.

The literature on conventional risk adjustment does not contain a formal statement of why conventional risk adjustment is the solution to an adverse selection problem in the market for health plans. No one, so far as we know, has supplied the argument for why the economic loss associated with a deviance between the risk adjusted payment and the predicted cost should go up according to the square of the difference. In this section, we follow the existing risk adjustment literature and simply presume that a better statistical fit of predicted to actual costs, as measured by an R^2 statistic, advances the cause of dealing with the selection problem related to individual access.

Before we go on to discuss the solution for MVORA, it may be worthwhile to present conventional risk adjustment in the context of our model. Conventional risk adjustment derives weights from the solution to an ordinary least squares regression of M on X.[7] Thus, conventional risk adjustment is, in fact, the solution to Eq. (12) without the constraints in Eq. (13) and is given by:

$$\beta^{c} = (X'X)^{-1}X'M$$

The conventional risk adjustment formula is therefore,

$$r_i^{c} = X_i\beta^{c}$$

We can express the set of S linear constraints in Eq. (13) in matrix form as $A\beta = C$, where A is an $S \times J$ matrix with elements

$$a_{sj} = \begin{cases} \dfrac{\operatorname{cov}(m_s,x_j)}{\bar{m}_s} - \dfrac{\operatorname{cov}(m_S,x_j)}{\bar{m}_S} & s = 1, \ldots, S-1 \\ \bar{x}_j & s = S \end{cases}$$

β is a $J \times 1$ vector of weights on the adjuster variables and C is an $S \times 1$ vector of constants with element

$$C_s = \begin{cases} \dfrac{\operatorname{cov}(m_s,M)}{\bar{m}_s} - \dfrac{\operatorname{cov}(m_S,M)}{\bar{m}_S} & s = 1, \ldots, S-1 \\ b & s = S \end{cases}$$

Note that the solution for MVORA described above is the estimated coefficients from a regression of M on X (i.e. as in conventional risk adjustment), constrained by the S linear restrictions in Eqs. (13) and (14) (Theil, 1961). Thus, we can write MVORA as

$$\beta^* = \beta^{c} + (X'X)^{-1}A'[A(X'X)^{-1}A']^{-1}(C - A\beta^{c}) \tag{15}$$

One can see that if the constraints in Eq. (13) are binding, MVORA will be different than conventional risk adjusters. One special case in which the constraints do not bind is when conventional risk adjustment is 'perfect' in the sense of completely capturing the variance in health care costs. In this case, $\Sigma_j \beta_j x_j = M$, and conditions (13) are exactly satisfied. Except in this extreme case, the constraints (13) will be binding and MVORA will differ from conventional. We can also observe that adding a new risk adjuster variable x_j will improve the

[7] A simple regression of M on X ignores the distributional characteristics of M. For many observations, $M = 0$ (no health care spending). Furthermore, among the positive observations, M is skewed to the right. Most of the empirical literature studying the effect of economic factors on health care use has employed a two-part model originally developed for the RAND Health Insurance Experiment (Duan et al., 1983). In the first part, a use–no use equation is estimated with a logit or a probit regression. In the second part, an OLS regression is performed on a transformation (square root, log) of the dependent variable for the observations with positive spending. See Manning (1998) and Mullahy (1998) for recent discussion. The risk adjustment literature relies primarily on one-part linear regressions. The main rationale appears to be a desire for simplicity in the risk adjustment formula.

regulator's ability to minimize Eq. (12). Adding a new risk adjuster with at least some correlation with some element of cost gives the regulator another instrument and must do at least as well in minimizing Eq. (12).

As the presentation of MVORA makes clear, the risk adjustment formula we propose depends on plan profit maximization. Strictly speaking, MVORA is the minimum variance risk adjustment formula that sustains the efficient expenditure decisions by plans as profit maximization. The question arises as to how to figure MVORA when expenditures are simply 'data' from some pattern of service expenditures generated by some other process, for example a fee-for-service system in which there is no health plan paid by capitation, and in which profit maximization by a health plan plays no role.[8] Any change in the payment system, for example, introduction of a risk-adjusted capitated system, will change incentives and change the pattern of services. A MVORA system can be calculated with first and second moment information from the data from the FFS system, but then introduction of the payment system would change the use patterns. It seems likely that a recalculation of the MVORA formula with the shift in use patterns, and perhaps a further recalculation, etc., would lead towards optimal service patterns, but such a question of the dynamic adjustments to risk adjustment is not formally confronted here. We are making only a weaker claim, that if the health plan is providing the efficient level of services while profit maximizing, MVORA will keep it there. (The conventional risk adjustment scheme would move the plan away from optimality).

5. An example: conventional risk adjustment and MVORA

Like conventional risk adjustment, MVORA can be found by using regression techniques in data. In this section we calculate conventional risk adjustment and MVORA in a very simple empirical example, to illustrate how MVORA works, and to show how it differs from conventional in the way premiums are figured.

Suppose data on per person spending comes from the following simple environment. There are two services a and d. The population is divided into three age groups, age 1, age 2 and age 3. The proportion of the population in age j, is α_j. Age is the only available risk adjuster. For each individual i, $x_{ij} = 1$ if the individual is in age j, and $x_{ij} = 0$ otherwise.

If an individual needs service a, the dollar value of services he will receive is m_a, regardless of his age group, and if an individual needs service d the dollar value of services he will receive is m_d, regardless of his age group. The proportion of individuals who need service a is the same in all three age groups and

[8] Medicare's risk adjustment system to pay HMOs is, for example, calibrated on data from the FFS sector.

normalized to 1 and the proportion of individuals who need service *d* in age *j* is γ_j, $j = 1,2,3$.

In figuring MVORA, the information assumption we make in this example is the following. There are two types of people, the healthy and the sick, differentiated by their use of service *d*. The healthy do not need *d* at all, the sick always need *d*. We assume that people know their type with certainty. That is, healthy people can accurately forecast that they will use m_a of service *a* and none of service *d*, and the sick type accurately forecast that they will use m_a of service *a* and m_d of service *d*. So, γ_j is the share of the sick types in age group *j*.

The age categories will be correlated with type so long as the γ_j differs for each age category. Age will thus have value as a risk adjustor. Conventional risk adjustment will be unable to fully solve the quality problem since under conventional risk adjustment private information remains about type and a health plan would want to 'underprovide' service *d* and 'overprovide' service *a* to attempt to attract the healthy within each age category. MVORA fixes this problem.

The calculations below are simplified because the age categories are (0,1) and the anticipated use of m_d by the healthy type is zero. This allows us to avoid explicit summations over types in the covariance terms below.

Let r_j^c denote the conventionally risk adjusted premium paid for an individual in group *j*, then:

$$r_j^c = m_a + \gamma_j m_d \tag{16}$$

Conventional risk adjustment simply pays for each individual the average cost of an individual in the age group the person belongs to.

Moving to optimal risk adjustment, we need to define some covariances for this example. One should first notice that since all individuals in all age groups use the same level of service *a*,

$$\text{cov}(m_a, x_j) = 0, \quad \text{for } j = 1,2,3 \tag{17}$$

and

$$\text{cov}(m_a, M) = 0, \tag{18}$$

where *M* denotes the total dollar value of services the individual receives.

As for service *d*, under our assumptions:

$$\text{cov}(m_d, x_j) = \alpha_j(\bar{m}_{dj} - \bar{m}_d) \text{ for } j = 1,2,3 \tag{19}$$

where:

$$\bar{m}_{dj} = \gamma_j m_d \tag{20}$$

is the average dollar value of services *d* used by an individual in age *j*, and

$$\bar{m}_d = \sum_{j=1}^{3} \alpha_j \bar{m}_{dj} \tag{21}$$

is the population average use of service d.

Furthermore, since m_a is the same for all individuals, we get that:

$$\mathrm{cov}(m_d, M) = \mathrm{var}(m_d) \tag{22}$$

In order to calculate the optimally risk adjusted premium we assume that the plan breaks even namely,

$$b = m_a + \bar{m}_d \tag{23}$$

where b is the budget per person.

Note that under the conventionally risk adjusted premium above the plan also breaks even, so the comparison between the two payments schemes will be on an equal basis.

Plugging Eqs. (17)–(23) into Eqs. (10) and (11) we obtain that optimal risk adjustment is a triple $(\beta_1^{\mathrm{o}}, \beta_2^{\mathrm{o}}, \beta_3^{\mathrm{o}})$ that solves the following two equations:

$$\sum_{j=1}^{3} \beta_j^{\mathrm{o}} \alpha_j (\bar{m}_{dj} - \bar{m}_d) - \mathrm{Var}(m_d) = 0 \tag{24}$$

and

$$\sum_{j=1}^{3} \beta_j^{\mathrm{o}} \alpha_j = m_a + \bar{m}_d \tag{25}$$

The MVORA is the triple $(\beta_1^*, \beta_2^*, \beta_3^*)$ that solves

$$\underset{\beta_1, \beta_2, \beta_3}{\mathrm{Min}} \sum_{j=1}^{3} \alpha_j [\gamma_j (m_a + m_d - \beta_j)^2 + (1 - \gamma_j)(m_a - \beta_j)^2] \tag{26}$$

s.t. Eqs. (24) and (25).

Table 1 presents a special case of the example above.

The first column in Table 1 specifies the three age groups, the second specifies

Table 1
An example of Conventional and Minimum Variance Optimal Risk Adjustment

Age group	α	γ	m_a	m_d	r^c	r^*
1	0.6	0.05	150	200	160	5.28
2	0.2	0.2	150	200	190	325.37
3	0.2	0.3	150	200	210	538.77

α, the proportion of each group in the population, the third column specifies γ, sickness' d probability, the fourth column specifies m_a, the dollar value of service a, the fifth column specifies m_d, the dollar value of service d. The last two columns compare r^c, the conventionally risk adjustment premium, with r^*, the MVORA premium, for each age group.

As can be seen from the table, the conventional and MVORA premiums are quite different. For the relatively healthy age group 1 (the 'young') MVORA pays much less than conventional risk adjustment whereas for the other two age groups it pays much more. As is well-known, since age is an imperfect indicator of type, a conventionally risk adjusted premium will leave a health plan with some incentives to attract the healthy types (those that would use only service a) within each age group. The conventional premiums would therefore induce a plan to 'oversupply' treatment for service a and 'undersupply' treatment for service d. MVORA premiums counteract this incentive. By undersupplying service d, the plan would tend to discourage relatively more of the age 3 people. By paying more for the age 3 group, MVORA gives the plan a strong incentive to keep them. Indeed, MVORA figures the risk-adjusted weights so that the incentive to supply both service a and d is just balanced to efficiently allocate a fixed budget. The marked disparity between the conventional risk adjustment and MVORA is a consequence of the extreme assumption we make about private information. If individuals could not perfectly forecast their health care use, MVORA would be closer to conventional.

One can also return to Eq. (5) above to examine shadow prices if plans were paid by conventional risk adjustment. In our extreme example, the situation would be disastrous for service d. Since everyone who has positive spending on service d is a loser (negative π), the denominator of Eq. (5) is so small it is negative! Essentially, a plan has an incentive with conventional risk adjustment to reduce service d to as low a level as possible.

6. Uncertainty and asymmetric information

The analysis so far has assumed that both the plan and the individual know with certainty the spending in dollars on each of the health care services that will be used by the individual, m_{is}. We shall now extend our model to the more realistic case where both the individual and the plan only hold some beliefs about the future health care needs of the individual, and these beliefs may not the same. We will see that the two crucial elements in the analysis here are the plan's beliefs about each individual i's expected use of each service s, which we denote by $\tilde{m}_{is}$, and the plan's beliefs about individual i's beliefs of how much of each of the services he is expected to use, denoted by $\hat{m}_{is}$.

In Appendix A, we analyze plan's behavior in this case and show that the profit maximizing shadow prices are given by:

$$q_s = \frac{\sum_i n_i \tilde{m}_{is}}{\sum_i \Phi_i' \, \hat{m}_{is}(r_i - \tilde{M}_i)}. \tag{5'}$$

A plan's expected profit for an individual is the product of the probability the person joins the plan times the profit or loss once he is a member. The probability that he joins depends on what the person believes he will receive in terms of services in the plan. The profit once he is enrolled depends on revenue and use in the plan. It is the plan's expectations about these two elements that will govern plan behavior in rationing care. Note that with respect to the decision to join a plan, the plan must form beliefs about what the consumers' expect. This is $\hat{m}$. Once in the plan, and from the point of view of the plan, what consumers expect is no longer relevant – what matters is what the plan expects a person to use. This is $\tilde{m}$. The placement of $\hat{m}$ and $\tilde{m}$ in Eq. (5′) reflects these considerations. See Appendix A for derivation.

Now MVORA can be redefined to be the solution to the following problem:

$$\underset{\beta_1, \ldots, \beta_J}{\text{Minimize}} \ \frac{1}{N}\sum_i \Big(\tilde{M}_i - \sum_j \beta_j x_{ij}\Big)^2 \tag{12'}$$

$$\text{s.t} \quad \sum_j \beta_j \left[\frac{\text{cov}(\hat{m}_s, x_j)}{\bar{m}_s} - \frac{\text{cov}(\hat{m}_S, x_j)}{\bar{m}_S}\right] = \frac{\text{cov}(\hat{m}_s, \tilde{M})}{\bar{m}_s} - \frac{\text{cov}(\hat{m}_S, \tilde{M})}{\bar{m}_S}$$

$$s = 1, \ldots, S-1 \tag{13'}$$

$$\text{and} \quad \sum_j \beta_j \bar{x}_j = b \tag{14'}$$

The objective function in Eq. (12′) is a modification of Eq. (12) to take into consideration *plan's beliefs* about its expected loss (profit) on each individual. Condition (12′) is meant to capture the problem of individual access. As such, it is plan's expectations about what persons will cost in the plan that govern plans' decisions about access. We believe it is reasonable to suppose that if $\tilde{M}_i$ differs from r_i, access problems may emerge. Therefore, minimization of the sum of the squares of the deviations of r_i from plan's expectations of costs, $\tilde{M}_i$, is a reasonable criterion. The constraint (13′) comes from Eq. (5′) in a way similar to the way we have obtained Eq. (13) from Eq. (5).

The solution for MVORA described above is the estimated coefficients from a regression of $\tilde{M}$ on X, constrained by the S linear restrictions. Thus, we can write MVORA as

$$\beta^* = \tilde{\beta} + (X'X)^{-1}A'[A(X'X)^{-1}A']^{-1}(C - A\tilde{\beta}) \tag{15'}$$

where, $\tilde{\beta} = (X'X)^{-1}X'\tilde{M}$ and, A is an $S \times J$ matrix with elements

$$a_{sj} = \begin{cases} \dfrac{\text{cov}(m_s, x_j)}{\bar{m}_s} - \dfrac{\text{cov}(m_S, x_j)}{\bar{m}_S} & s = 1, \ldots, S-1 \\ \bar{x}_j & s = S \end{cases}$$

β is a $J \times 1$ vector of weights on the adjuster variables and C is an $S \times 1$ vector of constants with element

$$C_s = \begin{cases} \dfrac{\text{cov}(m_s, M)}{\bar{m}_s} - \dfrac{\text{cov}(m_S, M)}{\bar{m}_S} & s = 1, \ldots, S-1 \\ b & s = S \end{cases}$$

In the case of uncertainty and asymmetric information the conventional risk adjusters will be different than MVORA for two reasons. First, even if the constraints (13′) were not binding, the solution to MVORA will be $\tilde{\beta}$ and not β^c. Second, if the constraints are binding, the solution will be given by the constraints and will not come from an unconstrained regression.

7. Discussion

The quality of health care offered by competing managed health care plans is the foremost concern of health policy in the U.S. and many other countries. Many elements of quality cannot be controlled by regulation, leaving open the opportunity for plans to manipulate the quality they offer in an effort to achieve a profitable mix of enrollees. Another policy concern is that plans may take actions to discriminate against particular individuals. The method of risk adjustment proposed here contends with both the quality and the access problem in an empirically implementable way. Conditions for optimal risk adjustment can be derived from the means and covariances among risk adjuster variables and elements of health care spending. A statistical fit criterion can then readily be applied to select among the weights satisfying conditions for efficient quality.

Application of our methodology requires that the distribution of health care spending in a population be available at the level of the service, not just in total. The concept of a service represents the level of aggregation at which management of a health plan makes decisions about resource allocation. In most health care claims data sets, information on the diagnosis, location of services (office, hospital, etc.), procedure conducted, and specialty or type of provider is typically included on the claim. Information is certainly available to classify health care encounters into the ‘services’ called for in our paper; the issue is, however, how all this data should be used to do so. This requires new work on ‘classification’ of health care use from the point of view of resource management. Theoretical work can consider the implications of using ‘too fine’ or ‘too coarse’ methods of aggregation for purposes of deriving conditions for optimal risk adjustment.

Our method also calls for information about expected health care costs, specifically, plans expectations about the consumers' costs, and plans beliefs about consumers' expectations. While this may seem a dismaying prospect, in this respect we are no worse off than conventional risk adjustment. Careful discussions about the goals of conventional risk adjustment have recognized for some time that the objective of conventional risk adjustment is to explain 'predictable costs,' yet it is actual costs that we see in the regressions deriving recommended weights. The same fall back of using the distribution of actual spending as opposed to expected spending is available to our method as well.

Another way to look at the issue is as a topic for research. Health plans make the decisions about resource allocation and access on the basis of expectations. Regulators attempting to counter the adverse consequences of these decisions would benefit by knowing how plans make decisions and based on what information. The saliency of plans' expectations to the policy issues related to quality and access makes this a very important topic for empirical research.

Acknowledgements

This paper was written while McGuire was at Boston University. Research support from grants R01 MH59254 from the National Institute of Mental Health and P01 HS10803 from the Agency for Health Research and Quality are gratefully acknowledged. We are grateful to Margarita Alegria, Ana Balsa, Zhun Cao, David Cutler, Randy Ellis, Richard Frank, Alberto Holly, Mathias Kifmann, Kevin Lang, Joseph Newhouse, Mark Satterthwaite, Manuel Trajtenberg, and Alan Zaslavsky for comments on an earlier draft. Remaining errors are the authors' responsibility.

Appendix A. Profit maximization with uncertainty and asymmetric information

Let T denote the set of possible health states of each individual and let t denote an element in T. Let $v_t = \{v_{t1}(m_{t1}), v_{t2}(m_{t2}), \ldots, v_{ts}(m_{ts})\}$ denote the vector of S valuation functions for the S services, if an individual's health state is realized to be t. We assume that for each t and s, $v_{ts}(\cdot)$ satisfies the properties discussed earlier.

Let $\check{x}$ be some random variable, the value of which depends on the state t, and let k be a distribution function defined over T. Let $E_k[\check{x}]$ denote the expected value of $\check{x}$ with respect to the distribution k.

The order of moves is as follows: at the first stage, the plan chooses its level of shadow prices $q = (q_1, q_2, \ldots, q_S)$. When choosing q, the plan is uncertain about each individual i's health state t, and it holds prior beliefs about it, denoted by (the distribution over T) g_i. At the second stage, individuals observe the plan's level of shadow prices and decide whether or not to join. At this stage, the individual is uncertain about his true health state t and he holds some prior beliefs about it. Let f_i be a distribution function over T which denotes the *plan's beliefs about individual i's beliefs* about his health state t. Thus, both g_i and f_i are plan's beliefs, the first is the plan's beliefs about individual i's health state and the second is its beliefs about what the individual thinks his health state to be.

At the third stage, when services are provided, the individual's 'true' health state is already known. Hence, for a given shadow price q_s and a valuation function v_{ts}, the plan's expenditures on this individual in service s will be $m_{ts}(q_s)$, given by:

$$v'_{ts}(m_{ts}(q_s)) = q_s.$$

Let $v_t(q) = \Sigma_s \; v_{ts}(m_{ts}(q_s))$.

Let $\bar{u}_t$ denote the individual's utility if his health state is t and he chooses the alternative plan. The plan's assigned probability that the individual will join is given by (we drop momentarily the subscript i from the analysis):

$$n_f(q) = 1 - \Phi(E_f[\bar{u}_t - \check{v}_t])$$

The plan's expected profit on the individual is

$$\pi_{fg}(q) = n_f(q)\left(r - E_g\left[\sum_s \check{m}_{ts}(q_s)\right]\right).$$

Differentiating with respect to $q_{s'}$ yields

$$\frac{\mathrm{d}\pi_{fg}(q)}{\mathrm{d}q_{s'}} = \Phi' E_f[\check{v}'_{ts'}\check{m}'_{ts'}]\left(r - E_g\left[\sum_s \check{m}_{ts}(q_s)]\right]\right) - n_f E_g[\check{m}'_{ts'}]$$

Using $v'_{ts} = q_s$ and $m'_{ts} = (e_s m_{ts})/q_s$ for all t, where e_s is the elasticity of demand (common to all individuals) for service s (see FGM, 1999), the right-hand side becomes

$$e_{s'}\left(\Phi' \hat{m}_{s'}(r - \bar{M}) - \frac{n_f \tilde{m}_{s'}}{q_{s'}}\right)$$

where $\hat{m}_{s'} = E_f[\check{m}_{ts'}]$ is the plan's belief about the individuals' prediction of the health resources he will consume, $\bar{m} = E_g[\check{m}_{ts'}]$ is the plan's prediction of its expenditure on the individual, and $\bar{M} = \Sigma_s \; \bar{m}_s$.

With a population of N individuals, the profit-maximizing q_s will therefore be (5′) from text.[9]

[9] Inducing a plan to supply services requires some profit. In the case of a perfect fit between a risk-adjusted payment r_i and $\bar{M}_i$, expected profit on each person would be zero. In this case, some fixed costs are necessary (as in any model of monopolistic competition).

References

Chapman, J.D., 1997. Biased Enrollment and Risk Adjustment for Health Plans, unpublished PhD dissertation, Harvard University.

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

Duan, N. et al., 1983. A comparison of alternative models for the demand for medical care. Journal of Economic and Business Statistics 1, 115–126.

Ellis, R.P., Pope, G., Iezzoni, L. et al., 1996. Diagnosis-based risk adjustment for medicare capitation payments. Health Care Financing Review 17 (3), 101–128.

Encinosa, W., 1999. The Economic Theory of Risk Adjusting Capitation Rates, Agency for Health Research and Quality, unpublished.

Enthoven, A.C., 1993. The history and principles of managed competition. Health Affairs 12 (Supplement), 24–48.

Frank, R., Glazer, J., McGuire, T., 2000. Measuring adverse selection in managed health care. Journal of Health Economics 19, 829–854.

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

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

Keenan, P., Beeuwkes, B.M., McGuire, T., Newhouse, J., 2001. The prevalence of formal risk adjustment, Fall, Inquiry, Harvard University.

Ma, C.-T.A., 1994. Health care payment systems: cost and quality incentives. Journal of Economics and Management Strategy 3 (1), 93–112.

Manning, W.G., 1998. The logged dependent variable heteroskedasticity and the retransformation problem. Journal of Health Economics 17 (3), 283–296.

Miller, R.H., Luft, H.S., 1997. Does managed care lead to better or worse quality of care? Health Affairs 16 (5), 7–25.

Mullahy, J., 1998. Much ado about two: reconsidering retransformation and the two-part model in health econometrics. Journal of Health Economics 17 (3), 247–282.

Neudeck, W., Podczeck, K., 1996. Adverse selection and regulation in health insurance markets. Journal of Health Economics 15, 387–408.

Newhouse, J., 1996. Reimbursing health plans and health providers: efficiency in production versus selection. Journal of Economic Literature 34, 1236–1263.

Rogerson, W., 1994. Choice of treatment incentives by a nonprofit hospital under prospective pricing. Journal of Economics and Management Strategy 3 (1), 7–51.

Selden, T., 1999. Premium subsidies for health insurance: excessive coverage versus adverse selection. Journal of Health Economics 18, 709–725.

Shen, Y., Ellis, R.P., 2000. Cost Minimizing Risk Adjustment, unpublished, Boston University.

Theil, H., 1961. Economic Forecasts and Policy, 2nd Edition. North-Holland.

van de Ven, W.P.M.M., Ellis, R.P., 2000. Risk adjustment in competitive health plan markets. In: Culyer, A.J., Newhouse, J.P. (Eds.), Handbook of Health Economics. North-Holland.

Weiner, J., Dobson, A., Maxwell, S. et al., 1996. Risk adjusted capitation rates using ambulatory and inpatient diagnoses. Health Care Financing Review 17 (3), 77–99.