Annu. Rev. Public Health. 1995. 16:401–30

RISK ADJUSTING COMMUNITY RATED HEALTH PLAN PREMIUMS: A Survey of Risk Assessment Literature and Policy Applications

Mita Giacomini
Center for Health Economics and Policy Analysis, Department of Clinical Epidemiology and Biostatistics, HSC-3H25, McMaster University, Hamilton, Ontario L8N 3Z5, Canada

Harold S. Luft
Institute for Health Policy Studies, University of California, San Francisco, California 94109

James C. Robinson
School of Public Health, University of California, Berkeley, California 94720

KEY WORDS: capitation fee, insurance selection bias, risk factors, reimbursement, actuarial analysis

ABSTRACT

This paper surveys recent health care reform debates and empirical evidence regarding the potential role for risk adjusters in addressing the problem of competitive risk segmentation under capitated financing. We discuss features of health plan markets affecting risk selection, methodological considerations in measuring it, and alternative approaches to financial correction for risk differentials. The appropriate approach to assessing risk differences between health plans depends upon the nature of market risk selection allowed under a given reform scenario. Because per capita costs depend on a health plan's population risk, efficiency, and quality of service, risk adjustment will most

0163-7525/95/0510-0401$05.00

strongly promote efficiency in environments with commensurately strong incentives for quality care.

INTRODUCTION

Capitated prospective financing is an increasingly popular approach to controlling health care expenditures. Capitation fixes health plan or provider revenues at levels based on the average anticipated health resource needs of members of a population, with the expectation that the program will supply necessary care prudently under the fixed budget. This principle is fundamental to community rated health plan premiums, but also applies to other health care financing mechanisms such as global budgeting at a regional level and capitated payments to physicians. The recipient of capitated payment faces a strong incentive to maintain costs below the prespecified reimbursement level. This can be accomplished in three ways: (*a*) by providing services more *efficiently*, either fewer in number or less resource intensive by unit for a given health outcome; (*b*) by undertreating selected individuals or *lowering quality* of services overall; or (*c*) by serving a population with health care requirements below the expected needs on which the payment rate was set, through *risk selection*.

Prospective payment is intended to motivate the first of these three organizational responses, improved efficiency. However, it must be applied within a context that inhibits the two alternative strategies of poor quality care and risk selection. Good quality measures and control mechanisms remain in the early stages of development; their improvement will become increasingly crucial to purchasers responding to health plan performance under competitive, capitated conditions. This paper focuses on policy approaches to redressing the second "pathological" response to capitation, risk selection. Risk adjustment corrects financially for risk selection, to make it a less profitable strategy than improving efficiency, given adequate quality control.

Health plans with sicker enrollees must charge higher premiums than plans with healthier members to cover their costs. *Adverse selection*, or enrolling a disproportionate share of unhealthy members, makes plans less price competitive to potential customers. Eventually it leads to plans either avoiding sick persons or offering them high premiums, which can become unaffordable. Plans successful at enrolling low-risk members should not be rewarded with market advantage for doing so. The purpose of risk adjustment is to reallocate revenues, or alternatively, from low-risk plans to high-risk plans. In either case, the redistribution goes only to the extent that the high-risk plans are equitably compensated for assuming more than their fair share of the sick in the market. Risk adjustment of community rated premiums financially pools

underlying population health risk across all plans in a market. Ideally, it also leaves the competing plans financially at risk for their own efficient operation, as well as the insurance of random or preventable changes in members' health.

The purpose of this paper is to highlight the issues that arise in applying the evolving empirical tool of risk assessment in the context of market reforms. Risk assessment and adjustment have promise, as well as important contextual and technical limitations, for combating the deleterious effects of biased selection among competing health plans. This paper reviews risk assessment models for application to risk adjusting community-rated health plan premiums under managed competition through a survey of the empirical literature and recent policy debates regarding the potential role for risk adjusters. First we present theories regarding the nature of risk selection in competitive health plan markets and introduce basic principles for adjusting capitated rates to correct for risk differences among competing plans. We then survey the empirical literature on risk assessment, which is the statistical method for measuring biased selection. We close with a critical analysis of how risk assessment approaches relate to risk adjustment and other market or regulatory controls on risk selection.

BIASED SELECTION IN HEALTH PLAN MARKETS

Risk selection occurs as the result of various actions by consumers, employers, and health plans, which can actively or passively segment risks. Risk segmentation can be the result of active efforts by health plans to attract low-risk enrollees or avoid high-risk enrollees; it can be the result of passive factors such as the location of facilities or the actions of other health plans; or it may result from the level and structure of employer-paid health benefits. The mechanisms and effects vary depending on the regulatory and competitive structure of the market. Unfortunately, most empirical literature to date on risk selection has treated the *criteria* for choosing a particular plan as a black box. Studies instead have focused primarily on which types of consumers pick which general type of plan (usually HMO vs FFS). Relatively few studies have ventured further to examine the role of generic plan features in repelling or attracting particular risks.

Discussions of biased risk selection provide a number of credible (although largely unproven) hypotheses about both plan and market features that might affect biased selection. First, a number of factors could determine whether a particular plan itself is susceptible to adverse selection. Health plans may affect their risk through a number of mechanisms; these are summarized in Table 1. Currently, health plans can select risks on two levels: through the employers or groups they contract with, and through the subscribers within each group who might select the plan over others offered (23). A plan may design and

package its product specifically to attract low risks and discourage high risks (19). High-quality benefits for the typical complaints of the healthy (e.g. maternity care, sports medicine, short-term psychotherapy) and spottier benefits for chronic health care conditions (e.g. restricted access to specialists, maximum coverage limits, etc) tend to attract better risks (17, 22). Bonus preventive services may appeal more to the healthy than to the ill (22). Inconvenient care for chronic diseases might be used to discourage low risks (17, 25, 27). Targeted advertising can appeal to healthy lifestyles, and enrollment procedures can require education or mobility correlated with better health (17). Sicker persons tend to be more attracted to low-cost sharing in exchange for high but predictable premiums, wider selection of physicians, and access to prestigious specialists. They are more likely to be in older plans (typically, but not necessarily, non-HMOs) because of long-term relationships with their plan's contracting providers (18, 19); these relationships make them less willing to switch plans (12, 22, 38). Because sick persons visit physicians more frequently, plans with geographically centralized providers may appeal to a smaller proportion of the sick in a region than plans with well-dispersed providers.

Alternatively, the plan may explicitly try to exclude high risks, for example through underwriting for preexisting conditions or risk factors. Plans may avoid epidemiologically risky communities by contracting only with providers in certain geographic locations (i.e. wealthier and healthier neighborhoods) or by redlining (or refusing to sell to) certain neighborhoods (e.g. gay communities to

Table 1 Health plan attributes potentially affecting biased selection

Mechanisms	Brings in higher risks (Adverse selection)	Brings in lower risks (Favorable selection)
Copayment/premium tradeoff	Low copayment/high premium	High copayment/low premium
Covered services	More comprehensive Conventional medicine . . .	Less comprehensive . . . with preventive emphasis
Geographic location of contracting providers	Less healthy, poorer areas	Healthier, wealthier areas
Geographic area covered by providers	Broadly distributed within region	Concentrated within region
Specialists, tertiary care	High quality, accessible	Low quality, inaccessible
Age of plan	Older, established plan	Newer, less familiar plan
Choice of providers	Free choice of provider	Restricted networks
Medical screening, underwriting	Prohibited	Allowed
	No preexisting exclusion	Exclusion of preexisting conditions
Rating method	Community rating	Experience rating

avoid persons with HIV) (22). Geographically concentrated providers may selectively inconvenience those frequently ill who are dispersed through a region (23). Finally, plans that historically have differentiated their prices across employer groups on the basis of risk or past claims (e.g. indemnity plans) may be less likely to have established relationships with high-risk employer groups (22).

A number of market and regulatory features also potentially affect plans' more passive susceptibility to selection resulting from information asymmetries between plans and consumers, sometimes referred to as consumer self-selection. Various health care reform proposals may modify these features to control the degree of risk segmentation, or to achieve other policy aims. First, some reforms aim at mandated universal coverage, whereas others aim more modestly at universal opportunities to buy coverage. The latter type of reform particularly invites adverse selection by persons who buy health insurance only on the occasion of becoming ill. Guaranteed issue of coverage and requirements for plans to cover preexisting conditions would further compound consumers' incentives to do so; together such reforms could result in higher premiums across the market if the healthy can avoid subsidizing the sick by staying out of the health insurance market entirely. Open enrollment and lock-in periods control the frequency with which consumers may change plans; shorter periods between plan choices generally encourage risk segmentation as consumers respond opportunistically to changes in their own health (22, 29). One-year lock-in, for example, may lower the incidence of risk-motivated plan switching compared to the one-month HMO lock-in currently used by the Medicare Risk program (22). Even longer lock-in periods could further decrease risk selection as fixed memberships' risk differences regress to the mean (37, 38). Requiring purchasing alliances or employer groups to offer their members a number of diverse plan choices also naturally increases opportunities for consumers to self-select according to risk, but encouraging continuity of coverage, rather than tying coverage options to the vagaries of employment options, may reduce selection.

Employer contribution strategies also affect employee price sensitivity and consequent self-selection among plans; reforms requiring employers or alliances to contribute only a flat amount approximating the lowest-cost plan toward premiums would raise price consciousness the most (6, 10, 21, 28). Because healthy persons are more price sensitive, more risk selection might ensue. Employers who pay a large percentage of all employees' premiums have good reason to encourage high utilizers to select the low-cost plan. Employers may also resist joining group purchasing arrangements (e.g. health alliances) if such pooling with alliance members might increase their premiums; plans in turn may avoid high-risk members by selectively contracting with low-risk employers or private, limited alliances. Some risk selection will also inevitably occur at random; this factor may impair the market entry and

survival of smaller plans, and could inhibit efficient pricing in all plans. Random risk fluctuations may become less important if the market consolidates into a few large plans under the increased competitive pressures and risk is spread across larger numbers within plans. The evolution over time of markets under proposed reforms will also affect likely risk segmentation patterns. Early in the process of implementing managed competition reforms, consumers may suddenly face more choices among health plans with standard benefits, and there could be large swings in enrollment. Traditionally, the more price-sensitive low risks move to less expensive plans, but their price sensitivity may be largely a function of willingness to switch providers, not of illness per se (5). If high-risk persons are able to change health plans without changing providers, their price sensitivity may increase. This may be expected in metropolitan markets where plans have overlapping provider networks, as well as wherever point-of-service options are mandated for HMOs. Stability of the supply side of the market will also affect risk selection. There is a "regression to the mean" effect as members age in their chosen plan (37), so fewer occasions of plan entry and exit from the market could mean more moderate biased selection. Finally, if the reform ensures universal coverage, an influx of price-sensitive, low-income persons who had been previously uninsured or publicly insured (and largely underserved) could also affect risk distributions.

Many risk-selection mechanisms may be amenable to regulatory control. Reform proposals commonly limit opportunities for health plans to select risks actively, by requiring a standard package of covered services, limiting cost-sharing options, prohibiting selective underwriting practices, requiring fair and uniform advertising practices, enforcing quality standards in care for the chronically ill, and prescribing community rating of premiums. Nevertheless, risk selection could not be eliminated entirely. To the extent that a reform allows health plans to differentiate their products *at all* on the basis of cost sharing, price, or contracting providers, they will attract different risks. Risk selection would also continue because employers and consumers have their own motives and methods for self-selection, and these practices are generally less appropriate to regulate because they are the same ones that drive market choices on the basis of efficiency, price, and quality. It is unlikely that any set of reforms will eliminate risk differences among health plans, so specific methods to assess and adjust for remaining differences will be necessary.

RISK AND RISK ADJUSTMENT

Risk adjustment policy is based on the simple idea that epidemiological factors partly determine health care utilization, which in turn partly determines health care expenditures. Given standardized benefits, variations in expenditures between health plans which are not due to these underlying health needs could

be attributed to efficiency differences in care delivery or administration. Risk, for risk-adjustment purposes, is a population's innate need for and propensity to use health care, independent of utilization (in)efficiencies. It is operationalized as essential health care expenditures. Hornbrook & Goodman (14) accordingly have defined risk as, "... the expected value of the distribution of per capita costs of efficiently provided preventive, diagnostic, and therapeutic health care services delivered to a defined group of enrollees for a specific future period."

Risk assessment models use a set of independent variables, risk factors, or risk adjusters, to predict necessary expenditures. For a study population, each risk factor is statistically associated with the dollar amount by which it typically increases or decreases the expected annual health care costs of an individual. Usually multiple regression or analysis of variance techniques are used. The resulting parameters, or cost weights, can then be applied to the members of another group to predict its total expected costs, or absolute risk. The group's risk is estimated by summing the expected expenditures of its members as predicted by the model. Relative risks are the ratios of these estimated total costs (absolute risks) between groups. To generate standardized relative risks, several groups are compared to a standard average-risk group, usually the potential market. For example, a health plan with a standardized relative risk of 1.1 would have 10% above average-risk members, compared to other plans serving the same market. It might legitimately require revenue 10% higher than the revenues of competing plans, due to its case mix and not its relative inefficiency.

There are two policy goals of risk adjustment: (*a*) to allow consumers to compare premium price differences that are not distorted by the health risk differences between the plans' memberships, but rather vary with the plans' value and efficiency, and (*b*) to reimburse each plan fairly for the proportion of population health risk that the plan assumes. The population is simply the market of all potential consumers who are choosing between a set of competing plans. Successful risk-adjustment policy means that if an efficient plan happens to enroll the sickest people in the market, it will be still be able to market at a competitive premium without this low premium translating into inadequate revenues to care for its needier members.

The financial transfers based on risk differences between plans may be approached in a number of ways. The first is to adjust community-rated premiums according to plans' relative risks. If plans rate prospectively on the basis of the market (e.g. alliance) population, then consumer prices automatically would be risk neutral, but plan revenues would have to be adjusted for risk segmentation. Higher-risk plans would receive more per capita than their quoted premium, and lower-risk plans would receive less. Alternatively, if plans adopt the community rate based on expected costs of current member-

ship, premium prices will be distorted by the plans' relative risk advantages or disadvantages in the market. Low-risk plans with lower per member per month expenses will naturally be able to offer lower rates. In this case, risk adjustment of market prices is required to raise the effective consumer price of low-risk plans, and lower that of high-risk plans. This may be done simply by adjusting the enrollee's premium contribution. If risk-selection patterns do not change after open enrollment, plans may be paid fairly at their quoted (not marketed) rates. If further risk segmentation changes as a result of open enrollment, a second adjustment of revenues might be in order.

There are alternatives to this standard approach to risk adjusting community-rated premiums. In a less price competitive environment, a large payer might use risk assessment information to formulate fair fixed rates that all health plans must accept as payment. This process is analogous to the "Adjusted Average Per Capita Cost" (AAPCC) concept used by the Health Care Financing Administration (HCFA) to determine reimbursement rates for Medicare risk contracting. The formula may rely on absolute risk calculations (actual expected costs produced by a risk assessment model), or apply relative risk estimates to modify some standard premium.

An alternative to community rating is to allow health plans to quote different rates for different risk categories. This can undermine the social insurance objective of reforms mandating community rating, but, like risk-adjusted community rating, helps enhance price competition based on efficiency rather than risk selection. A version of this method is being implemented in California's new purchasing pool for small groups. The risk-specific rates can be charged directly to members, such that high-risk consumers pay more than low-risk, or they can be composited by the purchasing group(s) to have low-risk members subsidize high-risk members. The latter method is somewhat analogous to community rating by class (CRC) practiced by some HMOs, except that the health plan quotes and receives class-specific payments rather than the group's average-risk class rate. Another key difference is that if the rate is composited, the purchasing group becomes financially at risk for added premium costs if high-risk members tend to choose less efficient plans. Unlike CRC, the individual plan is not vulnerable to inadequate payment if it experiences adverse risk selection from that particular purchasing group.

A final alternative is the high-cost condition pool, a form of prospectively priced reimbursement for unusually expensive, clinically specific conditions or condition-treatment pairs. The purpose is to reimburse plans fairly for disproportionate adverse selection by extreme cost outlier cases, to the extent the high costs are due to inevitable health care needs and not extravagant overutilization (classic stop-loss reinsurance can indiscriminately reward overtreatment as well, and thus deter efficiency). This method can be used to supplement other types of risk adjusted community rating and has recently

been implemented in New York State's small group purchasing pool. The condition pool approach estimates expected expenditures for selected health problems and pays plans either a lump or capitated monthly sum based directly on this. In the New York model, reimbursements are set at slightly below expected cost to create a disincentive to draw on the pool. Conditions covered in this manner include transplants for certain end-stage organ diseases, very low birthweight, advanced HIV disease, and ventilator dependency for ALS, severe trauma, or muscular dystrophy (33).

RISK ASSESSMENT

The following discussion of risk assessment modeling focuses on the problem of adjusting comprehensive community-rated premiums according to relative risk differences between competing health plans. However, many of the technical and conceptual issues generalize to all applications of risk information to determining appropriate prospective reimbursement rates.

The Dependent Variable: Health Expenditures

There are several ways to define the risk assessment model's dependent variable, future per capita health care expenditures. The most common is to use the dollar claims from a fee-for-service (FFS) plan. Studies based on group practice model HMO populations simulate FFS charges by imputing resource costs (11, 13, 15, 16, 35). Actual claims based on providers' comparable charges to non-HMO customers may be available for network model HMOs. Imputation is required in data systems that only measure utilization, rather than expenditures or resource use. It typically involves applying an average value (absolute or relative) to particular services, such as physician office visits or hospital admissions. Imputed costs tend to vary less than actual claims, and the restricted range can erroneously inflate the explanatory power of the model.

Another problem is the notorious skewness of per capita health care expenditures. Most people incur low annual expenditures while very few incur extremely high expenses. Linear risk assessment models can be quite sensitive to outliers (11, 16); models using log transformations or multiple equations provide a more accurate fit (26, 27). Estimating annual costs for persons who die or disenroll from a group during the year can compound the skewness problem. For example, when the costs of a "million dollar baby" who dies after a three-month life under neonatal intensive care are annualized, the result is not only an outlier, but an outlier of a magnitude beyond financial possibility. On the other hand, removing persons who die from the analysis introduces bias because mortality is systematically correlated with both risk factors and expenditures. Researchers often log or truncate expenditures, and sometimes use mortality or part-year enrollment as a control variable or selection criterion

to deal with these problems. However, while investigators must often use such methods to deal with limited data and analytic constraints, health plans are liable for real, not log expenditures and cannot ignore outlier cases.

The dependent variable becomes further complicated when health care is conceptualized as more than medical care. This issue arises, for example, in comparative risk analyses of Medicare risk contracting, where HMO enrollees receive more preventive and supplementary services at low out-of-pocket cost than do their FFS counterparts. Similarly, some large employers argue that reforms should not force them to pool risk with other groups, because their health promotion programs actively invest in employee health and so they should reap the rewards in terms of lower medical insurance expenses. Benefits such as covered services and cost-sharing arrangements must be well defined, and preferably standardized between any populations whose risk is being compared. Supplemental care and spending on health promotion ideally should be included in a risk assessment model if it has potential impact on externalized health care costs or benefits, utilization of covered services, and longer term health risk. Whether payers, patients, or others should be able to recapture the difference in expected costs attributable to their health-promoting or risk-taking behavior is a question for another paper.

A critical and difficult risk-modeling task involves differentiating efficient and necessary care from superfluous care. Efficient means having a high ratio of benefits to costs. Quality of care, supplier-induced demand, moral hazard, and nonmedical aspects of consumer demand all confound risk assessment's assumption that historical correlates of expenditures can be used as meaningful proxies for health care needs. Risk assessment models can be judged in part by how cleverly they control for the effects of costs potentially containable through efficiency improvements, and focus instead on those driven by epidemiological and demand characteristics beyond the control of the plans or the providers.

The Independent Variables: Risk Factors

Risk assessment models are characterized by their independent variables, or risk factors. The major variables are: (*a*) those epidemiologically associated with populations' morbidity or demand patterns (demographic factors); (*b*) those that are more direct proxies for individuals' health conditions (health status factors); and (*c*) those that measure clinical precursors of health problems (clinical factors). Table 2 offers a fairly exhaustive list of the types of variables that have been proposed as risk adjusters. Several features determine whether a particular variable is appropriate to use for risk assessment: its conceptual relationship to health risk, the ability to measure it accurately from available data, its statistical contribution to accurate predictions of expense, its social acceptability for application to health care financing, and its susceptibility to gaming by financially interested parties.

Table 2 Variables proposed or used for risk assessment models based on individual or family unit of analysis

Independent variables	
Demographics, Socioeconomic Status, or Local Market Characteristics	
Age	Job classification
Gender	Education level
Family size	Geographic location
Family composition	Industry of employment
Marital status	Institutional status
Ethnicity	Transportation access
Primary language	Supply of providers, facilities
Welfare status	Urban residence
Welfare eligibility	Familiarity with services
Poverty	Supplementary insurance coverage
Body weight relative to height	
Income	
Employment status	
Health status	
Self-reported health status	Diagnosis based on prior use of:
Mental health	Hospital services
Prior expenditures	Outpatient services
Quantity of inpatient services	Prescription drugs
Quantity of outpatient services	Use of ancillary/support services
Quantity of drugs	Disability status
Mortality (population rates)	Functional impairment
Health risk	
Clinical values:	Behaviors predisposing disease:
Laboratory values	Social support
Genetic screening	Nutrition, weight
Physical exam findings	Smoking
Dependent variables	
Claims based, e.g. FFS claims	
Resource value based, e.g. HMO costs imputed from encounter data	
Total health care costs, including consumer out-of-pocket costs	
Costs for covered benefits only	

Risk factors should relate closely to the health status of individuals or to the epidemiology of communities. Demographics such as age, gender, ethnicity, marital status, and family size correspond to population determinants of health relating to genetics, aging, and fertility, as well as to economic status and social support. Welfare status and employment data can be proxies for socioeconomic status and lifestyle factors associated with health risk. Some demographic risk factors pertain to demand characteristics as well as epide-

miology, for example, geographic proximity of providers, familiarity with services, and transportation access. There is some controversy as to whether these demand or taste factors are appropriate risk adjusters. If their effect on utilization were truly exogenous to the delivery system (that is, members could not be convinced to utilize services more efficiently), they would be acceptable adjusters (14).

Direct measures of health status may come from a variety of sources. The most common are utilization data. However, all measures of utilization tend to be confounded to varying degrees by practice style or efficiency differences among the plans. Prior health expenditures and crude utilization data (e.g. quantity or incidence of services) are most susceptible to this bias. Utilization data that yield specific morbidity information by diagnosis allow for discriminating risk factors on the basis of whether they are chronic or acute, severe or not, and in some cases, more amenable to managed care or not. Mortality rates may be useful predictors of risks in groups. Surveys that question people directly about their health bypass confounding interactions with the health care system. However, like demographic factors, perceptions can reflect culture or taste factors as much as physical health determinants of demand.

In addition to conceptual appropriateness, accurate measurements of the variables should be available both to generate good estimates and to deter gaming. For adjustment of per capita premium rates, the adjuster should be measurable on an individual or family unit level. This is most easily done through administrative databases generated for other purposes in the health care market, such as medical records, claims, or personnel files. It may also be possible to collect data specifically for risk assessment (e.g. through special population surveys) if the information will improve risk prediction substantially over more easily available data and if it is worth the additional cost. Age and gender are available from almost all potential data sources. Personnel file variables such as job classification, marital status, family composition, salary, and education can be used to improve demographic models. Administrative files maintained by public agencies also yield useful risk information; Medicare, for example, uses age, gender, welfare status, institutional status, and residence county expenditure levels to determine the AAPCC that it pays HMOs under risk contracts. Finally, medical utilization and clinical data contained in automated medical records can be used for risk assessment based on direct health status measures. Databases vary in accuracy and in quality of information. Data used for financial purposes (e.g. claims) tend to be audited more carefully than data used for other administrative or bureaucratic purposes (e.g. encounters).

The risk adjuster must contribute substantially and significantly to explaining expenditures. This contribution is not an immutable statistical fact, but will be affected by the particular incentives, opportunities, and information driving

the risk-selection process. Statistical power may be sacrificed to meet other criteria for a good risk assessment model. For example, limiting past hospital diagnoses used to predict next year's costs to those that are relatively non-discretionary mitigates providers' incentives to hospitalize inefficiently, but also decreases statistical power because many meaningful diagnoses cannot be used (2).

Regardless of how accessible, measurable, or powerful a predictor may be, other social values might interfere with its measurement or application. For instance, privacy considerations may limit access to information such as individuals' income, self-reported health, functioning, mental health, genetics, ethnicity, or health-related behaviors. Potentially conflicting interests on the part of the organization administering the risk assessment, for example, government agencies or employers, naturally compound privacy concerns. Some critics also worry that explicitly associating cost factors with risk characteristics such as race, language, disability, or economic class might reinforce social stigmas. Conversely, however, these risk factors could improve medical access for disadvantaged groups if they result in higher financial rewards for members' enrollment or treatment. The Clinton Health Security Act, for example, specified risk adjusting premiums according to socioeconomic status and mental health (although these factors have been little studied as risk assessment variables).

Finally, measurements of the adjuster should not be easily gameable, that is, distorted by market participants trying to enhance their profit from the financial transfers based on the assessed risk differences. Gaming could occur in a number of ways: Data could be falsified or, probably more likely, assignment of individuals to ambiguous risk categories could be distorted. This is most likely with utilization-based health measures (e.g. diagnoses), and less feasible with self-reported health status or demographic statistics. An example has been seen in the case of DRG creep under prospective payment for hospital care: Patterns of disease incidence appeared to shift as patients were classified into higher-profit diagnosis categories (7, 8, 32). Finally, utilization-based measures of health status such as diagnosis at hospitalization can create incentives for providers to overutilize services on which lucrative flags are based.

Study Population

The study population used to estimate the parameters can determine how well the cost weights generalize to other populations. Artifacts of cost data, such as imputing incomplete costs or truncating very high costs, can affect the estimated impact of risk factors on expenditures. The study group's benefit coverage will also influence the pattern of expenditures. Although differing benefit packages can be partially controlled for by removing services from the analysis to simulate a standard package, the influence of omitted services on

the use of covered services is lost. Dual coverage and supplementary benefits patterns can also distort a risk assessment model because (*a*) some costs may not appear in the data set, and (*b*) the cost-containing impact of copayment is lessened. Mortality and migration (between health plans as well as regions) introduce sampling and measurement problems.

The association of risk factors with individuals' expected expenditures may also be biased wherever unmeasured population-specific social and epidemiological determinants of need interact with the risk factors included in the model. For example, a risk assessment model may yield different cost weights for age and gender if it is estimated on data from a privately insured population than it might on data from a Medicaid-insured population.

Different risk factors predict well in different populations. In the elderly, for example, hospitalizations are both more frequent and more likely to be for chronic conditions; their measurement contributes important information to risk assessment. In the frail elderly, physical functioning may predict resource needs better than clinical diagnoses, which become numerous and interactive in very old age (M Hornbrook, personal communication). In contrast, the nonelderly are rarely hospitalized; when they are, it is most typically for acute conditions that do not affect next year's resource needs (15, 27). Consequently, hospital-based diagnoses are less useful predictors in this population; demographics and family composition may be better proxies for health care needs. Identification of the relatively few young individuals with chronic illnesses would of course improve any model. Indicators based on ambulatory diagnoses or sentinel drug prescriptions may detect more high-risk diagnoses than hospital discharge diagnoses would in the nonelderly.

Problems of empirical bias are compounded when generalizing risk models from one population to another. The best solution is to calibrate the model on a study population as representative as possible of the actual population across which financial risk adjustments will be made. For example, if risk adjustments will be made between health plan enrollees in a regional alliance, the model should be calibrated on a regionally representative population. Even so, there will be difficulty generalizing any model calibrated under today's financing environment to populations experiencing profound improvements in coverage and access under universal coverage and health care reforms. Risk forecasts for the uninsured, underinsured, or Medicaid-insured may be particularly underestimated owing to pent-up demand.

Accuracy

A number of empirical studies of potential risk assessment models have been published in recent years. For this review, we selected only studies using risk factors to predict or explain population *expenditures* for health care (and not merely service utilization). The study designs for these 17 papers are summa-

rized in the Appendix. Six research teams studied Medicare populations; 11 studied nonelderly populations. Five studies addressed HMO populations, and the rest used subjects covered by variants of FFS plans. Most investigators formulated models using risk information from one year to predict the following year's expenses, but several modeled expenditures instead as a function of concurrent risk status. The risk factors in the latter models would appear more powerful than they would be if applied to prospective financing. The remainder of our review focuses on how the various risk factors appear to perform across this diverse body of literature.

Several measures are currently available for evaluating risk assessment models for accuracy. Each addresses somewhat different aspects of a model's performance. The most intuitive are measures of correlation. These represent how tightly actual expenditures match predicted expenditures, both at the individual and at the group levels. A related question is how far group predictions tend to be off on average, in dollar or relative terms. Bias is another issue: To the extent that a model predicts imperfectly, in what identifiable individuals or groups does it systematically overpredict or underpredict? This bias is especially important for subpopulations, such as the chronically ill, who could suffer discrimination as a result of cost underestimation.

The first criterion is how closely the model predicts expenditures at the level of the *unit of analysis* used to develop the model, i.e. the individual plan member or subscriber unit (subscriber and covered family members). For regression models, the R^2 statistic represents the fraction of the variation in expenditures *across individuals* that can be explained by the risk factors. This R^2 is useful for refining the statistical models; the higher the R^2, the better the model is at predicting individuals' health care costs. However, health care expenses for individuals are notoriously variable and unpredictable. A few studies have estimated that a risk assessment model accounting for all predictable expenditures might at best achieve an R^2 of only 0.15 for adults (26, 35) and 0.37 for children (27). This maximum R^2 statistic is specific to the population being studied; some published risk assessment models have achieved individual R^2s well above these limits (9, 16, 36).

Table 3 summarizes published risk assessment models calibrated on non-Medicare populations. It is important to note that these studies use varying population samples and analytic approaches. Because of this, the explained variance (e.g. R^2) of the models should not be compared directly between studies. In other words, a model showing a higher R^2 in one study is not necessarily better than a different model with a lower R^2 in another study. Within each class of adjusters, the relatively wide variation in R^2s demonstrates that model accuracy is clearly sensitive to a number of technical issues beyond whether *demographic* or *health status* adjusters are used: sampling approach, study population, modeling techniques (multi- or single-equation, functional

Table 3 Risk assessment models estimated on non-Medicare populations

Risk factors	Approximate R^2	Model	Population	1st author	Notes
DEMOGRAPHIC MODELS					
Age, gender	0.01	Predict	HMO Oregon adults	Hornbrook (13)	
Age, gender	0.02	Predict	HMO Oregon adults	Hornbrook (15, 16)	
Age, gender	0.03	Predict	HMO California adults	Hayes (11)	
Age, gender	0.04	Explain	HMO Minnesota adults	Weiner (36)	
AAPCC	0.02	Predict	FFS US adults	Newhouse (26)	4-equation model
Age, gender, site, AFDC	0.02	Predict	RAND except children	Newhouse (27)	2-equation model
Age, gender, employment variables	0.02	Predict	HMO Oregon adults	Hornbrook (16)	
Age, gender, supplement insurance, region	0.02	Predict	Insured Netherlands	van Vliet (35)	2-equation model
Age, gender, step	0.03	Explain	Bank FFS N California	Rosencranz*	Log costs
Demographics, socioeconomic variables	0.03	Predict	Insured Netherlands	van Vliet (35)	2-equation model
Age, gender, supplemental insurance, region	0.03	Predict	Random Netherlands	van Vliet (35)	4-equation model
Age, gender, step, employment variables	0.04	Explain	Bank FFS N California	Rosencranz*	Log costs
Demographics, socioeconomic variables	0.04	Predict	Random Netherlands	van Vliet (35)	4-equation model
CRC–age, gender, employment variables	0.21	Predict	HMO California	Cave (9)	Log costs
SELF-REPORTED HEALTH STATUS MODELS					
AAPCC, subjective health	0.03	Predict	FFS US adults	Newhouse (26)	4-equation model
Self-reported health status	0.04	Predict	HMO Oregon adults	Hornbrook (13)	
Demographics, subjective health	0.05	Predict	RAND experiment children	Newhouse (27)	2-equation model
Self-reported by parent	0.05	Predict	FFS US children	Newhouse (27)	2-equation model
Age, gender, self-reported health	0.06	Predict	HMO Oregon adults	Hornbrook (13)	
Self-reported health	0.18	Predict	HMO Oregon adults	Hornbrook (16)	
PHYSIOLOGICAL MEASURES					
AAPCC, physiological health (claims)	0.05	Predict	FFS US adults	Newhouse (26)	4-equation model
Demographic, physiological health	0.11	Predict	RAND experiment children	Newhouse (27)	2-equation model
Physiologic measures	0.11	Predict	FFS US children	Newhouse (27)	

PRIOR UTILIZATION					
Inpatient diagnosis	0.04	Predict	HMO California adults	Hayes (11)	
Prescription drugs	0.05	Predict	HMO Oregon	Hornbrook (15)	
Demographics, prior in/out use	0.07	Predict	Insured Netherlands	van Vliet (35)	2-equation model
Drugs, outpatient diagnosis	0.07	Predict	HMO Oregon	Hornbrook (15)	
ACGs–outpatient diagnosis cost groups	0.15	Explain	HMO Minnesota adults	Weiner (36)	
Age, gender, ADGs–outpatient diagnosis groups	0.19	Explain	HMO Minnesota adults	Weiner (36)	
Demographics, prior in/outpatient utilization	0.21	Predict	RAND except children	Newhouse (27)	2-equation model
PRIOR EXPENDITURES					
AAPCC, prior cost	0.06	Predict	FFS US adults	Newhouse (26)	4-equation model
Demographics, prior cost	0.07	Predict	Insured Netherlands	van Vliet (35)	2-equation model
MIXED MODELS					
Demographics, chronic condition, impair	0.08	Predict	Random Netherlands	van Vliet (35)	4-equation model
AAPCC, physiologic health, subjective health, prior use	0.09	Predict	FFS US adults	Newhouse (26)	4-equation model
Demographics, physiologic health, subjective health	0.11	Predict	RAND except children	Newhouse (27)	2-equation model
Demographics, chronic disease, functioning, self-reported health	0.11	Predict	Random Netherlands	van Vliet (35)	4-equation model
Demographics, physiological, subjective health, prior use	0.24	Predict	RAND except children	Newhouse (27)	2-equation model
In/outpatient utilization, expenditures	0.33	Predict	HMO California	Cave (9)	Log costs
CRC, prior in/outpatient utilization, expenditures	0.38	Predict	HMO California	Cave (9)	Log costs

Key: Predict = predictive model; Explain = explanatory models (contemporaneous)

*Rosencranz SL, Luft HS. 1993. Changing the focus: evaluating health care expenditure models on risk-stratified groups of enrollees. (Unpublished manuscript; 30a).

form, explanatory or predictive), data sources, measurement of the variables, and data modifications (cleaning, imputation of missing information, functional transformations, etc).

Simple age and gender models appear to explain 1–4% of cost variations among subscribers or individuals. Demographic models that add more detailed information regarding employment and welfare still explain only 2–4%. One model using *community rating by class* variables and HMO data explained an unusual 21%; imputed cost data and logarithmically transformed expenditures may have contributed to this unusual goodness-of-fit. More complex health status models usually include age and gender; these variables contribute independently to risk estimates and help minimize potential model bias and cross subsidization between old and young or male and female (15, 16). Self-reported health status measures in addition to age and gender perform slightly better, typically explaining 3–6% of the variance, and in one study up to 18%. Physiological measures based on physical examination explained 5–11%. Prior utilization models relying on the incidence, volume, or diagnosis recorded from either inpatient or outpatient episodes vary substantially in their predictive power, and explain 4–21% of individual differences in expenditures. Models incorporating diagnosis information seem to perform better than pure utilization models in adults. Prior utilization with demographics appeared relatively accurate in the one model calibrated on children. Individuals' prior health care costs plus simple demographics achieved relatively low R^2s of 6–7% in the two studies reviewing these types of models for comparative purposes. In contrast, another study combining prior spending with utilization information reported R^2s of over 0.3. Finally, models combining data from disparate sources (i.e. demographic, personnel, financial, medical records) reported R^2s ranging from 0.08 to 0.29. These models are probably the least administratively feasible to apply because of the extensive data required.

Table 4 surveys published risk assessment models calibrated on Medicare populations. As in non-Medicare populations, demographic AAPCC factors explained little of the cost variations between population members, from less than 1% to 3%. Models relying on direct measures of well being, such as disability, activities of daily living, or self-reported health status, achieved R^2s of 0.02 to 0.07. Similarly, prior utilization models not incorporating diagnosis information also yielded R^2s ranging from 0.02 to 0.07; utilization-based diagnosis models performed somewhat better, explaining 4–16% of the variance in expenditures. The two prior expenditure models explained 6–9% of individual variance. A model combining physiological measures of clinical risk, disability, and prior utilization in a small regional population produced an R^2 of 0.10.

In summary, models vary in their ability to predict (or in some cases, contemporaneously explain) expenditure differences between individuals. In

Table 4 Risk assessment models estimated on Medicare populations

Risk factors	Approximate R^2	Model	Population	1st author	Notes
DEMOGRAPHIC MODELS					
AAPCC	0.00	Predict	MC FFS Michigan	Lubitz (20)	
AAPCC	0.01	Predict	MC FFS National	Ash (3)	
Age, gender, welfare	0.01	Predict	MC FFS National	Beebe (4)	
AAPCC	0.03	Explain	MC S Carolina	Manton (24)	
AAPCC	0.03	Predict	MC Framingham	Schauffler (31)	Log costs
NON-UTILIZATION MEASURES OF HEALTH					
Perceived health status	0.02	Predict	MC FFS Michigan	Lubitz (20)	
IADL	0.03	Predict	MC FFS Michigan	Lubitz (20)	
PHYSIOLOGICAL MEASURES					
AAPCC, clinical risk factors	0.05	Predict	MC Framingham	Schauffler (31)	Log costs
AAPCC, disability	0.07	Predict	MC Framingham	Schauffler (31)	Log costs
PRIOR UTILIZATION MEASURES					
Demographics, any inpatient use	0.02	Predict	MC FFS National	Beebe (4)	
Demographics, any inpatient use	0.03	Predict	MC FFS National	Ash (3)	
Chronic conditions	0.04	Predict	MC FFS Michigan	Lubitz (20)	
Demographics, days inpatient	0.04	Predict	MC FFS National	Beebe (4)	
Demographics, DCG	0.05	Predict	MC FFS National	Ash (3)	
Demographics, inpatient days, any outpatient	0.05	Predict	MC FFS National	Ash (3)	
AAPCC, prior inpatient/outpatient use	0.06	Predict	MC Framingham	Schauffler (31)	Log costs
Prior inpatient/outpatient utilization	0.07	Predict	MC FFS Michigan	Lubitz (20)	
AAPCC, in/outpatient diagnosis	0.12	Explain	MC S Carolina	Manton (24)	
AAPCC, in/outpatient, mortality	0.16	Explain	MC S Carolina	Manton (24)	
PRIOR EXPENDITURES					
Prior expenditures	0.06	Predict	MC FFS Michigan	Lubitz (20)	
Part B expenditures only	0.09	Predict	MC FFS National	Ash (3)	
MIXED MODELS					
AAPCC, prior use, clinical risk	0.08	Predict	MC Framingham	Schauffler (31)	Log costs
AAPCC, prior use, clinical risk, disability	0.10	Predict	MC Framingham	Schauffler (31)	Log costs

Key: Predict = predictive model; Explain - explanatory models (contemporaneous)

addition, the wide variation in R^2s among models using similar variables illustrates the sensitivity of risk assessment model performance to research design. Some models produced high R^2s in part by using multiple or nonlinear equations. In general, demographic models, even fairly sophisticated versions, explain the least at the individual level. Not surprisingly, more direct measures of health status do better at predicting individuals' health care spending next year.

Ability to explain *individuals'* variations in expected costs is not the only measure of a good risk assessment model. The assessment of biased selection across health plans in a market requires that models be accurate when applied to *groups* of health plan enrollees; consequently, a fundamental issue is how well the model predicts expenditures within, and expenditure differences across, large groups. A statistic sometimes used to demonstrate this dimension of accuracy is the prediction error, or the related predictive ratio. Predictive error statistics compare predicted costs with actual costs (either the difference or the ratio) for selected groups or individuals. The mean prediction error is the average dollar amount by which estimates deviate from the real cost. To validate models, this statistic is commonly applied to random samples from the population on which the model was calibrated. Predictive errors appeal to policymakers because they can be used to assign a dollar value to a model's potential inaccuracy at the group or individual level. However, a major drawback of the statistic is that the dollar value is highly specific to the study population and may not be generalized with confidence to actual, risk-stratified groups. It is typically generated by estimating risks based on *random* samples, which, by definition, approximate average risk of the population from which they are drawn. Due to the law of large numbers, it is not surprising that, as larger samples are drawn, the predictive error improves; this relationship holds regardless of the how poorly a model's risk factors predict expenditures at the individual level. Even demographic models with relatively small individual level R^2s can yield highly accurate and reliable expenditure predictions when applied to large groups; demographic models, for instance, are highly accurate in randomly drawn groups of 1000 or more (30).

Most policy relevant are measures of a model's accuracy in groups that are *different from the sample* used to calibrate the model, and for comparing groups that *differ from each other* in their risk composition. Useful metrics include group-level prediction errors and group R^2. Group-level predictive error statistics have been used to test for bias in subpopulations of high- or low-risk individuals (rather than random samples) as defined either by the model's dependent or independent variables, such as women, persons with cancer or heart disease, persons with no prior costs, or those with a recent history of low or high hospital utilization (3, 4, 15, 16, 31). A group R^2 measures correlation between actual and expected costs across risk-segmented

groups. The group R^2 is generated by arranging a population into risk-stratified groups as defined by the expected value of the dependent variable (predicted cost), and then measuring the correlation between expected and actual expenditures by group. This measure has been used on fully risk-stratified groups (i.e. the 50 highest-risk persons in one group, the 50 next most expensive persons in the next group, and so forth) (SL Rosenkranz & HS Luft, unpublished manuscript). The simulations in an employed population yield a group R^2 of 58% for a simple demographic model and 63% for a model based on personnel data even though these models yielded R^2s of less than 0.04 at the subscriber unit level.

However, neither randomly selected nor fully risk-segmented groups occur in health care markets. One study has looked at naturally occurring risk selection, using 542 employer groups within a single health plan (11). Using the *group* as the unit of analysis, regression modeling yielded an R^2 of 0.51 for a demographic model, 0.52 for a model with inpatient admission data. In this study, the demographic model's predictive power at the *subscriber* level was approximately half that of a model incorporating hospitalization diagnoses, but not much less accurate at predicting actual costs of the 542 employer *groups*. Further research is needed to assess various models' performance under conditions of realistic risk segmentation, especially between health plans or providers rather than between employer groups within a single delivery system.

A risk assessment model should be unbiased, that is, it must not systematically under- or overestimate the expenditures of identifiable subpopulations (e.g. male vs female, young vs old, members of one plan vs another). The last criterion is particularly important: risk assessment models should not systematically reward low quality delivery systems by confusing the process or consequences of poor care with members' innate health risk. Models relying on prior utilization measures, for example, may favor *inefficient* health plans that indulge in more unnecessary utilization. Health surveys as well as utilization measures of diagnoses could favor *ineffective* health plans if bad health reflects a history of inadequate treatment or prevention in addition to exogenous influences (22).

Biases that erroneously favor particular demographic groups could create unfair financial subsidies (15, 16), as well as opportunities for competitive risk selection. The inclusion of a risk factor in the model (e.g. gender) tends to decrease bias (15), but unfortunately this does not necessarily ensure that the model will predict well for groups in one particular risk category (e.g. females). Higher predictive accuracy at the individual level has also been associated with lower age and gender bias at the group level (15). Wherever a model regularly overpredicts for one easily identified group and underpredicts for another, capitated health plans or providers can profit by pursuing the first and avoiding the second. The problem of bias directly relates to the problem of assessing goodness-of-fit at the level of group estimates under conditions of

strategic risk selection. A model that appears unbiased and accurate in random populations may be very biased or inaccurate in risk-segmented populations if the risk segmentation occurs on the basis of (or correlated with) risk factors that interact with other determinants in the model.

The Problem of Selection Within Risk Classes

Risk assessment models that appear powerful at the group level but weak at the individual level could, in theory, be defeated by sophisticated selection of risks on a case-by-case basis, using information not captured by the model (25). Health care spending is skewed and difficult to predict in general. Even within tightly defined risk categories (e.g. women aged 40–50 with cancer), expected expenditures will still vary widely between individuals. Because health plans and providers possess information that allows them to discern high-risk from low-risk individuals ***within*** most risk-adjuster classes, they could, in theory, game the system by catering to low risks and repelling high risks within defined risk classes. However, there could be several remedies to the potential problem of selective disenrollment.

Perhaps least effectively, sophisticated risk selection could be detected and corrected for with a model that is maximally accurate at the individual level. A perfect model would be prohibitively expensive to administer, and imperfect models fail to eliminate substantial profitability to health plans of aggressive disenrollment of high risks (26). Another approach involves enforcing high-quality standards or relying on professional norms to inhibit the plans' administrators or providers from actively discouraging targeted members. Probably the most compelling deterrent would be market reputation. Selective disenrollment means that one plan's sick members are systematically unloaded onto competitors (the Queen of Spades phenomenon in the game of Hearts). Especially under universal coverage where one plan's dump automatically becomes another plan's new member, any plan successful at selective disenrollment would be easily identified and exposed by its competitors. Unsatisfactory services for the chronically ill may also appear as lower quality and may be unattractive to healthy prospective customers, jeopardizing the plan's market share. In short, under competitive conditions it may be more profitable for plans and providers to care for their high-risk patients than to have their reputation—and market share—endangered by aggressive selective disenrollment strategies. Failing that, side payments from a patient's old plan to the newly selected plan may be a direct way of discouraging the practice (22).

DISCUSSION

Financially risk adjusting capitated premiums mitigate health plans' incentive to select risks, but this is not its primary purpose (22). Uneven risk between

health plans does not itself threaten effective competition. For example, it may be socially desirable for health plans to specialize in the efficient care of particular diseases or populations. Rather, risk adjustment reverses the undesirable results of risk segmentation, without seeking to undo all of the *processes* that create it. Adjustments must accomplish health risk pooling in spite of the emergence of risk-biased associations between consumers, employers, and health plans from a competitive market.

This raises two important points about designing risk-adjustment systems. First, risk adjustment achieves financial subsidization of the sick by the healthy. It supports a solidarity perspective of social health insurance rather than one based on actuarial fairness (34) because it requires all health plans in a market to contribute equitably to caring for the net burden of illness in the market population. The level at which risk is pooled, and consequently the means by which risk is adjusted, depend upon what subsidies are intended as well as allowed under health care reforms.

Second, the factors used for assessing risk must accurately detect the particular type of risk segmentation one expects to find in the reformed market. The risk factors must relate to the information asymmetries and strategic practices driving selection. As a rule, the less random and more opportunistic the risk selection, the more sophisticated the risk assessment model will need to be to estimate the risk differences correctly. For example, if consumers were randomly allocated to large plans, the resulting expected cost differences could be measured accurately by using very simple demographics. If, at the other extreme, plans were allowed to examine and exclude members on a case-by-case basis, risk assessment likewise must be based on more clinically detailed measures of group members' epidemiology.

As market reforms are implemented, any reasonable risk adjustment would be an improvement over unadjusted capitated financing. However, assessment methods and models will need to continually evolve in sophistication to keep up with strategic responses to reformed market conditions and risk-adjustment policy. There are several reasons for this. First, if reforms achieve their goal and all plans and providers become more efficient, the profitability of marginal improvements in efficiency will drop and the relative profitability of risk selection will rise. Second, with time and familiarity, payers and providers may develop new ways of gaming the adjustment process (25). More sensitive risk assessment models will be needed to detect potentially more devious risk selection. Third, part of the managed care philosophy is for the capitated health plan to take some responsibility for preventing disease and maintaining members' health. Although health care systems should not be penalized for the morbidity of cases they assume, they should not be rewarded for morbidity they cause, or fail to prevent. Dynamic risk adjustment models should be developed that appropriately consider risk as an output as well as an input of

health care. Quality management and risk adjustment can become a more continuous process if the health risk of inputs can be differentiated from the health risk of outputs. For example, in risk adjusting premiums between health plans, plans could be more generously rewarded for very sick persons who join than they are for members who become sicker with preventable diseases, or for very sick members who leave the plan.

Finally, risk assessment models will require regular re-estimation and calibration to keep pace with changes in epidemiology and medical practice. Technological improvements, clinical guidelines, and organizational efficiencies should be expected to—indeed, should be *designed to*—alter the relationship of risk factors to health care costs over time. If medical science and delivery systems truly progress over the years, high-risk categories should regress to the mean, while low-risk categories do not.

Health plan premiums are determined by a basic triad of risk, efficiency, and quality. Risk adjustment forces health plans to pursue efficiency and quality to achieve an attractive product and competitive pricing. Many discussions of risk adjustment policy focus on encouraging efficiency while leaving quality as a concern for other areas of health care reform. However, these three elements form a true system of incentives. Strong incentives toward one put pressure on the others, which in turn require commensurate support through appropriate regulatory and market incentives. Successful risk adjustment imposes burdens particularly on policy reforms in quality management. Pressures and means to demonstrate quality must rise in concert with disincentives to risk select in order to ensure that efficiency improvements rather than quality erosion become the most profitable enterprise for health plans.

APPENDIX Summary of empirical risk assessment studies surveyed

First author	Sample	N	Risk factors	Dependent variable
Anderson (1)	1983–84 FFS aged Medicare beneficiaries, 2.5% national sample	213,844	AAPCC CRCs: "Cost Related Groups" based on less discretionary inpatient diagnosis, demographics DCGs: "Diagnostic Cost Groups" based on less discretionary inpatient diagnosis + AAPCC variables PACs: "Payment Amount for Capitated Systems": inpatient MDC, chronicity, outpatient use, + demographics, disability	Parts A & B Medicare expenditures Truncated at 99th percentile
Ash (3)	1974–80 FFS aged Medicare beneficiaries; 5% national sample	18,677 for 1975–77 20,263 for 1978–80	AAPCC DCGs: "Diagnostic Cost Groups" based on less discretionary inpatient diagnosis + AAPCC variable	Medicare expenditures
Beebe (4)	1975–78 FFS aged Medicare beneficiaries; 5% national sample	20,773	Age, gender, Medicaid/welfare proxy Demographics + any inpatient Demographics + days inpatient	Total Medicare reimbursement
Cave (9)	1984–85 Staff HMO, families, ages 1–64 yrs, Southern California	24,330	CRC: "Community rating by class" age, gender, marital status, family size, family composition, industry Prior utilization and expenditures Combined CRC and prior expenditures	Inputed from HMO encounters, + contracting provider claims Log transformed

APPENDIX *(Continued)*

First author	Sample	N	Risk factors	Dependent variable
Hayes (11)	1984–88 PGP HMO subscriber units, subscriber under 65, employed, non-Medicaid, non-Medicare; Northern California	587,659	Age, gender, step Length of enrollment Nondiscretionary hospital admissions 2 yrs prior (consult w/MDs, 14 categories)	Imputed from HMO utilization records, no trans (truncated for comparison only)
Hornbrook (15)	1980–87 PGP HMO individual members under 65, employed, non-Medicaid, non-Medicare; Oregon	51,633 pooled time series cross section	Age, gender "Clinical-behavioral classes" of diagnoses (medical chart); Drug orders (medical chart) Prior year utilization expense	Imputed from HMO utilization records
Hornbrook (16)	1980–81; 1985–86 PGP HMO subscriber units, subscriber under 65, employed, non-Medicaid, non-Medicare; Oregon	1545 for 1980–81; 842 for 1985–86	Age, gender Education, occupation, marital status Self-reported health status Self-reported N outpatient visits Self-reported medical conditions	Imputed from HMO encounters
Hornbrook (13)	1990–91 PGP HMO individual members under 65, employed, non-Medicaid, non-Medicare; Oregon	8265	Age, gender Self-reported health status	Imputed from HMO encounters
Lubitz (20)	1982–83 FFS Medicare enrollees, Michigan	2000	AAPCC Perceived health status IADL Chronic conditions Prior expenditures Prior in/outpatient utilization	Medicare charges

Manton (24)	1981 FFS Medicare Southern Carolina	4000	AAPCC In/outpatient diagnosis Mortality	Medicare charges
Newhouse (26)	1974–82 individuals over age 14 participating in Rand Experiment	3958 (7690 person-years)	AAPCC: age, gender, site, AFDC status Physiologic health Subjective health Prior year inpatient/outpatient use	Claims
Newhouse (27)	Children *under* age 14 participating in Rand Experiment	1844 2185 person-years	AAPCC: age, gender, site, AFDC status Physiologic health Subjective health (parent's perception) Prior year inpatient/outpatient use	Claims
Robinson (30)	1981–84 FFS bank employee subscriber units; Northern California	31,849	Age, gender, personnel variables: step, educ, salary, occup level, marital, length employ, coverage eligibility full/part year	Log $ claims; truncated at $25K Part-year eligibility controlled with independent variables
Rosencranz*	1989 FFS bank employee subscriber units; Northern California	5000	Age, gender, step Marital status, salary, state of residence, length employ	Log $ claims Truncated at $50K
Schauffler (31)	1982–85 FFS aged Medicare beneficiaries; Framingham, Mass.	1162	AAPCC inpt/outpt use disability index clinical risk factors	Log $ payments

APPENDIX *(Continued)*

First author	Sample	N	Risk factors	Dependent variable
Van Vliet (35)	1976–80 individuals in large private insurance company, Netherlands	~35,000	Age, gender Supplementary insurance, geographic region	Claims
	1976 same as above	~14,000	Same as above, + family size, employment, income, education, urbanization Prior utilization Prior costs	Claims
	1981–82 individuals, national health survey data, Netherlands	~20,000	Age, gender, supplementary insurance, region Family size, "socioeconomic status:" body weight, urbanization, facilities supply Self-reported health status Chronic conditions, physical impairment	Imputed from utilization information
Wiener (36)	1980 or 1981; Network HMO individual members; Minnesota	Approx 35,000–40,000	Age, gender ADG—ambulatory diagnostic groups ACG—ambulatory cost groups	"Claims"—would be charges if FFS

* Rosencranz SL, Luft HS. 1993. Changing the focus: evaluating health care expenditure models on risk-stratified groups of enrollees. Unpublished manuscript. Earlier version presented at the annual meeting of the Association for Health Services Research, June 1993.

Literature Cited

1. Anderson GF, Steinberg EP, Powe NR, Antebi S, Whittle J, et al. 1990. Setting payment rates for capitated systems: a comparison of various alternatives. *Inquiry* 27:225–33
2. Ash A. 1994. Presented at Natl. Health Policy Forum, Washington, DC, Feb. 23
3. Ash A, Porell F, Gruenberg L, Sawitz E, Beiser A. 1989. Adjusting Medicare capitation payments using prior utilization data. *Health Care Fin. Rev.* 10:17–29
4. Beebe J, Lubitz J, Eggers P. 1985. Using prior utilization to determine payments for Medicare enrollees in health maintenance organizations. *Health Care Fin. Rev.* 6:27–38
5. Berki SE, Ashcraft MLF. 1980. HMO enrollment: who joins and why: a review of the literature. *Milbank Q.* 58:588–632
6. Bowen BD, Slavin EL. 1991. Adjusting contributions to address selection bias: three models for employers. *Adv. Health Econ. Health Serv. Res.* 12:77–96
7. Carter GM, Ginsburg PB. 1985. The Medicare case mix index increase, RAND (Santa Monica, CA), R-3292-HCFA
8. Carter GM, Newhouse JP, Relles DA. 1990. How much change in the case mix index is DRG creep?, RAND (Santa Monica, CA), R-3826-HCFA
9. Cave DG, Schweitzer SO, Lachenbruch PA. 1989. Adjusting employer group capitation premiums by community rating by class factors. *Med. Care* 27:887–99
10. Enthoven A. 1989. Effective management of competition in the FEHBP (Federal Employees Health Benefits Program). *Health Aff.* 8:33–50
11. Hayes ST. 1991. Demographic risk factors derived from HMO data. *Adv. Health Econ. Health Serv. Res.* 12:177–96
12. Hellinger FJ. 1987. Selection bias in health maintenance organizations: analysis of recent evidence. *Health Care Fin. Rev.* 9:55–63
13. Hornbrook MC, Goodman MJ. 1993. *Assessing relative health plan risk with the Rand-36 health survey.* Presented at Annu. Meet. Assoc. Health Serv. Res., June
14. Hornbrook MC, Goodman MJ. 1991. Health plan case mix: definition, measurement, and use. *Adv. Health Econ. Health Serv. Res.* 12:111–48
15. Hornbrook MC, Goodman MJ, Bennett MD. 1991. Assessing health plan case mix in employed population: ambulatory morbidity and prescribed drug models. *Adv. Health Econ. Health Serv. Res.* 12:197–232
16. Hornbrook MC, Goodman MJ, Bennett MD, Greenlick MR. 1991. Assessing health plan case mix in employed population: self-reported health status models. *Adv. Health Econ. Health Serv. Res.* 12:233–72
17. Jones SB. 1993. *Health plans, risk adjusters, and corporate alliances.* Testimony submitted to Subcomm. Health Environ., US House Represent. Comm. Energy Commer., Dec. 9
18. Jones SB. 1990. Multiple choice health insurance: the lessons and challenge to private insurers. *Inquiry* 27:161–66
19. Juba DA, Lave JR, Shaddy J. 1980. An analysis of the choice of health benefits plans. *Inquiry* 17:62–71
20. Lubitz J. 1987. Health status adjustments for Medicare capitation. *Inquiry* 24:362–75
21. Luft HS. 1986. Compensating for biased selection in health insurance. *Milbank Q.* 64:566–91
22. Luft HS, Miller RH. 1988. Patient selection in a competitive health care system. *Health Aff.* 7:97–119
23. Luft HS, Trauner JB, Maerki SC. 1985. Adverse selection in a large, multiple option health benefits program: a case study of the California Public Employees' Retirement System. *Adv. Health Econ. Health Serv. Res.* 6:197–229
24. Manton KG, Tolley HD, Vertrees JC. 1989. Controlling risk in capitation payment. Multivariate definitions of risk groups. *Med. Care* 27:259–72
25. Newhouse JP. 1994. Patients at risk: health reform and risk adjustment. *Health Aff.* 13:132–46
26. Newhouse JP, Manning WP, Keeler EB, Sloss EM. 1989. Adjusting capitation rates using objective health measures and prior utilization. *Health Care Fin. Rev.* 10:41–54

27. Newhouse JP, Manning WP, Keeler EB, Sloss EM. 1993. *Risk adjustment for a children's capitation rate.* Presented at Annu. Meet. Assoc. Health Serv. Res., Washington, DC, June 29
28. Robinson JC. 1993. A payment method for health insurance purchasing cooperatives. *Health Aff.* Suppl:65–75
29. Robinson JC, Gardner LB, Luft HS. 1991. Health plan switching in anticipation of increased medical care. *Med. Care* 31:43–51
30. Robinson JC, Luft HS, Gardner LB, Morrison EM. 1991. A method for risk adjusting employer contributions to competing health insurance plans. *Inquiry* 28:107–16
30a. Rosencranz SL, Luft HS. 1993. *Changing the focus: evaluating health care expenditure models on risk-stratified groups of enrollees.* Presented at Annu. Meet. Assoc. Health Serv. Res., Washington, DC, June
31. Schauffler HH, Howland J, Cobb J. 1992. Using chronic disease risk factors to adjust Medicare capitation payments. *Health Care Fin. Rev.* 14: 79–90
32. Simborg D. 1981. DRG Creep—A new hospital acquired disease. *New Engl. J. Med.* 304:1602–04
33. State of New York, Insurance Department. 1993. Regulation No. 146 (11 NYCRR 361): Establishment and Operation of Market Stabilization Mechanisms for Individual and Small Group Health Insurance and Medicare Supplement Insurance
34. Stone D. 1993. The struggle for the soul of health insurance. *J. Health Polit. Policy Law* 18:287–317
35. van Vliet RC, van de Ven WP. 1992. Towards a capitation formula for competing health insurers: an empirical analysis. *Soc. Sci. Med.* 34:1035–48
36. Weiner JP, Starfield BH, Steinwachs DM, Mumford LM. 1991. Development and application of a population-oriented measure of ambulatory care case-mix. *Med. Care* 29:452–472
37. Welch WP. 1985. Regression toward the mean in medical care costs, implications for biased selection in Health Maintenance Organization. *Med. Care* 23: 1234–41
38. Wilensky GR, Rossiter LF. 1986. Patient self-selection in HMOs. *Health Aff.* 5:66–80