

Paper prepared for Culyer, Anthony (ed.) **Encyclopedia of Health Economics**, Elsevier Press., Inc. 2014.

Risk Selection and Risk Adjustment

Randall P. Ellis and Timothy J. Layton

Boston University, Department of Economics

270 Bay State Road; Boston MA 02215 USA

ellisrp@bu.edu

Synopsis: Risk selection, which occurs when an individual's demand for a product is correlated with her risk, creates inefficiencies and inequalities in markets for those products and services. Studies have shown that risk selection often occurs in health care markets, especially in markets for health insurance. Risk adjustment is a method that has been developed to correct those inefficiencies by using models to calculate risk and compensate suppliers for the risk of each individual purchasing their products. These models have become more sophisticated in recent years and are currently used in the health care systems of a number of countries.

Keywords: Adverse Selection, Asymmetric Information, Health Insurance, Managed Care, Managed Competition, Medicare, Moral Hazard, Risk Adjustment, Risk Aversion

Introduction

The problem of risk-based sorting, often referred to as risk selection, and the use of risk adjustment to offset it are central concepts in health economics. After briefly defining risk selection and risk adjustment, this article provides an overview of the theoretical and empirical literatures that analyze these concepts. The issues covered here touch upon numerous entries in this book, including health insurance, adverse selection, health plan competition, death spirals, and quality report cards, among others (*Editors: Update with actual headings, which we cannot do easily*).

What is risk selection?

Risk selection occurs in health care markets whenever consumers differ in expected cost (risk) that cannot be priced and make choices based on differences in risk, shifting the risk from the individual to the supplier. This choice can result in potentially inefficient or unfair sorting by average cost, quantity of visits, or quality. The most common reason for unpriced variation in risk is asymmetric information, in which consumers have private information about their health status, environment, or tastes for health care that insurers are unable to use when setting premiums. Incentives for risk selection can also be created even with full information when pricing is regulated, such as when regulators restrict the information that health plans are allowed to use when setting premiums or benefit features.

What is risk adjustment?

While risk adjustment is defined in many ways, we offer one broad definition that includes almost all of the myriad ways the term is used: the use of information to explain variation in health care spending, resource utilization, and health outcomes over a fixed interval of time, such as a quarter or year. Although we do not discuss it here, the term risk adjustment is also used in the health services research literature to refer to methods of explaining variation in a particular procedure or episode of treatment.

Theory of Risk Selection

Almost all theories of risk selection center on the choice of health insurance plans. In the classic Rothschild and Stiglitz model of risk selection there are two types of consumers (high-risk and low-risk) and two states of the world (healthy and sick). The consumers differ in their probabilities of realizing the sick state, resulting in different expected costs in each potential state of the world. There is no moral hazard, so full insurance is optimal. Yet, under the assumptions of the model, a pooling equilibrium is either infeasible (when the low risk types are unwilling to purchase the plan priced at a pooled premium) or inefficient. This model is recreated in graphical form in Figure 1. The two axes measure available

consumption in each of the states of the world. If there is no insurance, available consumption is lower in the sick state due to health care spending; hence, the initial endowment with no insurance is at a point such as E for both consumers. Because the two consumers differ in the probabilities of the two states of the world, their indifference curves between different levels of spending will diverge, with low risk types having steeper indifference curves than high risks at every point. For risk-averse, utility-maximizing consumers, indifference curves between income in the two states of the world will be convex, and efficient consumption requires that insurance be provided until each type has equalized income in both states of the world (i.e., is on the 45 degree line). A possible outcome may be that the break-even pooled full insurance point in this framework is at a point such as C, which is preferred to no insurance by high risk types but less preferred than no insurance by low risks, making it infeasible and inefficient. Moreover, even if a sponsor (the government or an employer) forces this option to be offered, the health plan will strongly prefer enrolling the low-risk types and may either distort plan offerings so as to be less attractive to high risks (high deductibles or cost sharing) or take costly efforts to avoid high risks.

< Figure 1 near here >

While the Rothschild-Stiglitz model is quite nice for describing risk selection among traditional indemnity health insurance plans, selection in the real world is more complex. In the US, managed care organizations (MCOs) such as Health Maintenance Organizations (HMOs) and Preferred Provider Organizations (PPOs) have captured a large share of the market for health insurance from the traditional indemnity plans. These plans often offer much lower cost sharing in return for much more tightly rationed health care services. However, because consumer risk is still unpriced due to asymmetric information or regulation, the incentives for risk selection by profit-maximizing health plans are still quite strong, but the methods by which the selection occurs are likely to be quite different due to the low levels of cost-sharing that are a hallmark of MCOs. The classic model for describing risk selection among managed care plans was formulated by Glazer and McGuire.

The Glazer and McGuire model of risk selection moves away from selection on the level of cost-sharing offered by a plan toward a theory of “service-level selection.” Because high cost and low cost individuals demand different services, MCOs can induce the high cost individuals to avoid their plans by rationing the services demanded by these individuals more tightly than other plans. Likewise, they can attract low-cost individuals by rationing the services these individuals demand more loosely than other plans. For example, a profit-maximizing MCO may have incentives to ration care for a chronic condition like diabetes by including few or no diabetes specialists in its network. On the other hand, the MCO may want to provide easy access to acute services or alternative medicine like acupuncture or chiropractic services

in order to attract the low risks. In equilibrium, MCOs offer less than the efficient quantity of some services and more than the efficient quantity of others.

Both types of models of risk selection describe inefficient equilibria due to a correlation between demand and unpriced risk. These inefficiencies lead to welfare losses. Note that the correlation does not have to be positive (high risks demand more, or adverse selection) to induce a welfare loss. If risk and demand are negatively correlated (high risks demand less, or advantageous selection), welfare losses still occur, but now the losses are due to plans offering “too much” of something rather than too little. Also note that selection does not have to be limited to selection on cost-sharing or to service-level selection. Selection can occur on any attribute including health plan quality (or service-level quality), through special offers such as gym membership discounts, etc. The key result, however, is that when unpriced risk is correlated with demand, inefficiencies and welfare losses are likely to occur.

Empirical Models of Risk Selection

Empirical models of risk selection are important for two reasons. First, they help us to determine where selection exists and on what characteristics selection occurs. If we know where selection is a problem, we can implement solutions such as risk adjustment to fix the problem. If we know what characteristics selection occurs on, we can use regulation to limit its effects. Second, they allow us to measure the welfare losses from selection. With a measure of welfare loss, we can analyze trade-offs between inefficiencies caused by selection and inefficiencies caused by regulations intended to limit selection such as risk adjustment, reinsurance, and mandates.

A major difficulty involved with developing empirical models of risk selection is the confounding presence of moral hazard in health insurance. If there were no moral hazard, we could compare the average cost of individuals in Plan A with the average cost of individuals in Plan B and conclude that the plan with higher average cost is adversely selected. However, if Plan B has lower cost sharing or looser rationing of services and higher average cost, it is not clear whether the higher average cost is due to increased utilization due to the lower level of cost sharing (moral hazard) or to fundamentally higher cost individuals choosing Plan B because of the lower cost sharing (adverse selection). This problem can be solved with panel data and exogenous variation in health plan premiums, however. Essentially, moral hazard and adverse selection can be isolated by observing shifts in demand and corresponding shifts in average cost following a price change (see Einav and Finkelstein (2011) for a graphical description of this method). Further complexity arises if adverse selection and moral hazard interact with one another. Individuals may choose a plan with lower cost sharing because they have a higher elasticity of demand rather than due to their higher risk (Einav et al. 2013)

The method described above nicely allows for straightforward estimation of welfare losses. This method has produced estimates of welfare losses that are surprisingly small. These estimates are important because they allow researchers to use simulations to determine the welfare effects of various regulations such as incremental cost pricing, plan subsidies, or mandates. However, one strong assumption is necessary for the estimate of welfare loss to be complete: fixed contracts. The welfare losses being measured are really only those stemming from inefficient pricing of contracts. But the theoretical models of selection described above focus not just on pricing but also on the nature of the contracts themselves. The assumption of fixed contracts is likely valid in the context of employer provided insurance because the employer often chooses the plan parameters. However, in the context of a less regulated (or completely unregulated) market for insurance where insurers choose the majority of the parameters of the contracts they offer, the assumption of fixed contracts is likely to break down. In this unregulated environment, welfare losses occur not only through inefficient pricing of efficient contracts but through equilibria where only inefficient contracts are offered.

It is possible (and highly likely) that the welfare losses from distorted contracts are much larger than losses due to inefficient pricing. For example, coverage for mental health care in the US has been highly rationed in many health insurance plans because it attracts high risks. This is effectively a “death spiral” that has occurred in a plan characteristic rather than of an entire plan. When contracts are not fixed, it is important for empirical models to be able to highlight the characteristics that selection occurs on because, as the Glazer and McGuire model of service-level selection points out, these are the characteristics that will be inefficiently rationed, and, thus, these are the services that regulators must focus on in order to achieve efficiency and minimize welfare losses. However, empirical models that can quantify the welfare loss due to these inefficient contracts are few in number due to the fact that these contracts are extremely complex due to the seemingly infinite number of parameters firms can vary (network size, cost-sharing parameters, in-network hospitals, etc.). Nevertheless, there is empirical evidence that inefficiencies such as service-level selection by MCOs exists, just no easy way to determine how much welfare is lost due to these inefficiencies.

Instead of trying to quantify welfare losses due to inefficient contracts, the empirical literature has sought to answer the question of why the welfare losses due to inefficient pricing are so low. This has resulted in interesting new theories and empirical evidence for interactions between selection and market frictions such as imperfect competition and switching costs and behavioral issues such as inertia and other mistakes.

The discussion of risk selection so far has focused primarily on the US setting where numerous, diverse health plans compete in multiple dimensions (premiums, benefit features, cost sharing, and selective contracting) with an important goal of attracting profitable enrollees. Similar structures of competition exist in Chile and Colombia. Several other countries also have multiple competing health plans (e.g., Belgium, Germany, Japan, Netherlands, Switzerland, and Israel), however benefit features, cost sharing, and premiums are regulated much more tightly in these countries, and selective contracting is relatively rare. Although selection incentives exist in these other competing health plan settings, the plans typically do not control providers and have relatively few tools available for influencing selection. Selection problems are typically even less of an issue in countries with a single social insurance plan (e.g., Canada, France, Denmark, Italy, Sweden, and the UK), although selection issues can still arise through competition among individual providers or geographically, where consumers get to choose among alternative local market areas. Selection concerns are also common when there are private complementary or supplementary insurance policies alongside of a single, publicly-funded plan, as in Australia and Ireland.

Theory of Risk Adjustment

Remarkably, there is no unified or widely adopted theory of risk adjustment. Instead, there are models of risk selection that point to desirable features of risk adjustment models, and statistical models of risk adjustment that develop empirical risk adjustment models that satisfy various statistical properties (unbiasedness, minimum variance, robustness, and fair payment for subpopulations). One underlying reason why there is no unified theory of risk adjustment is that an important motivation for risk adjustment is usually equity, not just efficiency. With regard to efficiency-based arguments for risk adjustment, the appropriate risk adjustment model depends upon the market and regulations in which competing health plans (or providers) operate. If premiums, cost sharing, and benefit plans are allowed to vary across consumer attributes that are observable to the health plan, then there will be no unpriced variation in costs or selection problems, and only fairness and equity concerns will remain. Once regulators restrict premiums, cost sharing, and benefit coverage variation, risk adjustment is the classical tool for combatting risk selection.

Though there is no unified theory of risk adjustment, the literature has essentially assumed that the welfare loss from the distortions caused by selection are proportional to the sum of the squared differences between individuals' expected costs and the revenue a plan receives for those individuals. This assumption leads to the convenient result that the main goal of any risk adjustment system should be

to minimize this sum of squared differences, or to maximize the fit of the payment system as measured by the R-squared statistic. Other theoretical models of risk adjustment have built on this assumption.

Glazer and McGuire were the first to develop theoretical models characterizing “optimal risk adjustment,” which they distinguish from the existing statistical models that do “conventional risk adjustment.” The central objectives of conventional risk adjustment are unbiasedness (paying each plan so that predicted costs equal actual revenue for each individual) and maximizing predictiveness (minimizing deviations between payments and expected costs). The essence of optimal risk adjustment is to allow biased risk adjustment models which optimally correct for identified incentive problems in health care markets.

Glazer and McGuire (2000) choose to model the service distortion selection problem, in which competing health plans oversupply services that attract the healthy (e.g., acute care), and undersupply services that disproportionately attract the high cost, relatively sick (e.g., chronic care services). Since the signals used for risk adjustment are never perfect, even with conventional risk adjustment paying the expected costs it will be optimal for health plans to distort service offerings so as to attract the relatively healthy within a payment category, and deter the relatively sick. The solution Glazer and McGuire devise is to overpay on signals predicting a greater likelihood of being high cost, and underpay on signals predicting low cost, so as to undo the incentive to undertreat the high cost enrollees. For example if only half of patients with asthma in a plan have their diagnoses recorded in the base period, and the incremental cost of the observed asthma patients is \$500 higher than expected, then the plan should be paid twice this increment, or \$1000 to compensate the plan for the under-reported patients with asthma. Conversely, one should pay less than the observed average cost for healthy signals in order to keep overall payments neutral. This twist in payments can in theory undo incentives to undertreat in capitated payment systems.

The service distortion problem that Glazer and McGuire model is a particular problem in the US, since many plans use selective contracting to increase or reduce the availability of specific types of services or providers, thereby influencing the attractiveness of their plan. Similar incentives and concerns arise in other countries, such as Australia and Ireland, where private insurance plans are allowed to choose the extent of coverage for services or copayments not covered generously by the public system. Other selection problems can require different optimal risk adjustment adjustments. For instance, in the US and several European countries (Germany and Belgium) there are concerns about intentional distortion of the signals used for paying competing health plans, or “upcoding” the observed severity of patients.

Recent literature has explored risk adjustment in a setting where enrollee sorting on expected costs may not be as strong as sorting on the degree of risk aversion or other preferences. Two recent papers, (Bundorf et al. 2012; and Glazer and McGuire, 2011) introduce the possibility that the demand for

insurance is determined by both risk and taste and that there is not a perfect correlation between the two. This is especially relevant in the current environment in the United States where an integrated HMO may be able to provide care for a chronically ill patient at a lower cost than a PPO but the chronically ill may prefer the PPO due to its wider selection of providers. Both papers show that in this environment consumers will have different incremental marginal costs, but the only way to get consumers to sort efficiently across plans is to charge each consumer her particular incremental marginal cost. Thus, if premiums are uniform across individuals, individuals may not sort efficiently across plans, even with perfect risk adjustment. Glazer and McGuire examine the market equilibria that occur under different regulatory arrangements, analyzing the tradeoff between efficiency and fairness. They show that if taste can be used as the basis of payment, both efficiency and fairness can be achieved using a tax. However, when taste cannot be used as a basis of payment (because it is not observed) and health status must be used instead, subsidies and taxes based on health status are required to achieve both efficiency and fairness. In other words, a uniform payment along with perfect risk adjustment is not enough.

Recent theoretical work is beginning to examine how to implement risk adjustment in the presence of imperfect community rating, which is to say that insurance plan premiums are allowed to vary within specified limits across certain individual attributes (such as age and smoking status). Further work is also examining how risk adjustment models can accommodate intentional benefit plan variation, such as is being allowed in the US health insurance exchanges where substantial variation in cost sharing is being permitted. This theoretical work is important because, as explained above, the welfare loss from the distortions caused by selection incentives is proportional to the sum of squared differences between individuals' expected costs and the *total* revenues a plan received for those individuals. In the US health insurance exchanges the total revenues can come from multiple sources: premiums, risk adjustment transfers, reinsurance payments, and risk corridor payments. It is clear that these sources of payments will interact with each other and those interactions need to be identified in order to determine how well they will fix the problems of risk selection and what other distortions they may cause.

Empirical Risk Adjustment models

Early work in developing risk adjustment models focused on the statistical problem of maximizing the amount of variance in total spending that can be explained with available information (Ash et al, 1989; Newhouse et al., 1989). Even in this early work it was recognized that if lagged utilization or spending variables are used as explanatory variables, then the model is not only capturing the underlying illness burden, but also consumer taste for treatment, provider practice variation, or differences in the underlying efficiency of treatment, which may lead to incentive problems. European risk adjustment implementation

has been more precise than most US studies in distinguishing “acceptable costs”, viewed as appropriate for risk adjustment, and “unacceptable costs” which are viewed as ineligible for payment differentiation. Early work focused on self-reported measures from surveys that capture health status, however these measures are relatively expensive to gather and update, and not as highly predictive as insurance claims-based measures. In most modern risk adjustment models, diagnosis- and pharmacy-based information is used to predict spending. The extent to which each set of information is used in the models varies by country. The consensus view among risk adjusters and policy makers is that diagnoses and pharmacy signals, while not fully exogenous, are less endogenous than many other variables (such as health plan, provider type, access, taste, and consumer lifestyle), justifying their widespread use for risk adjustment.

From the onset, it has been recognized that health status information (whether self-reported, diagnoses or pharmacy) from the base period can either be used to predict outcomes from the same period or the subsequent period (i.e. the future). The former is called concurrent (or sometimes retrospective) risk adjustment, while the latter is called prospective risk adjustment, and the two vary only in the prediction period. Most payment systems use prospective risk adjustment, due to concerns about endogeneity of the signals as well as the practical reason that it means that risk factors on which payments are based can be measured a year earlier than the spending being predicted. Concurrent models always have higher explanatory power than prospective models. For quality measurement or normalization of many other performance or outcome measures, a concurrent framework is widely used. Careful comparisons of predictive power from the two frameworks are provided in a US Society of Actuaries study authored by Winkelman and Mehmud, and in a series of studies conducted at York University in the UK.

A useful early contribution in the risk adjustment literature used fixed effects in panel data to calculate a “lower bound on the upper bound” of what is potentially explainable at the individual level using time-invariant, prospective information. This method suggested that between 15 and 20 percent of the variance in spending was explainable using prospective variables. More recent studies suggest that the potentially achievable prospective R^2 is on the order of 25 to 35 percent of total health care spending and varies with the population, year, and data quality.

To illustrate the importance of using more information than just age and gender to predict costs, consider Figure 2, which plots average 2009 covered health care costs for each of 65 one-year age cohorts in the US Truven MarketScan commercially insured claims and encounter data. Figure 1 illustrates the importance of using relatively flexible specifications even for capturing age and sex adjustment of total health spending. The figure highlights that babies are disproportionately expensive, that women cost more than men through their childbearing years, and that in childhood males are slightly more expensive than

females. These patterns are poorly captured by including a linear age term or even when using third or fourth degree polynomials of age. Most sophisticated risk adjustment models calibrated on large samples use 30 or more age-sex dummy variables to capture this nonlinear pattern.

< Figure 2 near here >

As stated above, rather than only using (exogenous) age and gender, the most common approach used for risk adjustment is to use the rich information appearing on insurance claims as a proxy for individual health status. The most widespread information used is diagnoses, although pharmacy information is also common. Utilization measures (e.g., spending, hospitalizations, and counts of visits) are also highly predictive of future spending, although they contribute relatively modestly to the predictive power once a rich diagnostic model is used. Although claims-based information is only recorded when a visit to a health care provider is made, and is potentially “gameable” or amenable to manipulation, its strong predictive power and availability make it highly attractive.

Careful reviews of alternative risk adjustment models of total annual spending have been conducted in the US, Germany, and the UK, and are included in the recommended further readings section at the end of this entry. Table 1 contains a few highlights of five diagnosis-based risk adjustment models used for payment by public insurance programs in the US (Medicare and Medicaid), as well as large numbers of private health plans. The interested reader can view further details at the references noted in the table.

< Table 1 near here >

A glimpse at the dimensions along which many risk adjustment models vary is summarized in Table 2 from a Dixon et al. (2011) study using UK data. Looking first across the rows, age and gender alone only explain about 3-5 percent of total variation in spending at the individual level. Once diagnostic and prior utilization information are included in model (b), surprisingly little further variation is explained by including geographic variation (as captured by 152 geographical primary care trust dummies), 135 need variables (e.g., income, education, and prevalence of selected chronic conditions in the area) and 63 supply side variables (e.g., numbers of providers of various types and distances). Explanatory power at the individual level as measured by the R-squared differs only in the third or fourth decimal. The final row reveals that dropping the four prior utilization variables has a more significant effect on the model’s predictive power, reducing the model’s explanatory power by about half. Many would argue that the four lagged utilization variables are not only picking up health status heterogeneity, but also patient and provider taste variation. (Key “need” and supply side variables are still included in the model.)

Looking across the columns of Table 2 reveals that with 5 million observations in the estimation sample, there is no overfitting problem, even with over 500 right hand side explanatory variables. The final column shows that despite having only modest explanatory power at the individual level, where there is a great deal of individual patient randomness, the models do enormously better at the practice level where much of this randomness averages out. The third column sums up patient actual and predicted spending to the level of 797 primary care practices (averaging 6,500 patients per practice) before using the conventional R-squared formula to calculate predictive power. The explained variation in spending at the practice level starts at 34% for the age-gender model, and increases to just over 80% once geographic dummies are added in. Even the final model, which does not use the four utilization variables capturing patient and provider taste variation, explains 77% of practice-level variation in spending.

< Table 2 near here >

Risk adjustment has been used for over three decades for the US Medicare Advantage (Part C) program, which offers diverse, competing private health plans to elderly and disabled individuals in the US as a voluntary alternative to conventional Medicare. The risk adjusted payments to health plans from 1985 to 1999 used only age, gender, Medicaid eligibility, institutional status (i.e., whether in a nursing home) and the county of residence of the enrollee to determine the payment amount. Since 2000 risk adjustment in the US Medicare program has used diagnostic information, initially using only inpatient diagnoses, but since 2004 diagnoses from outpatient clinician claims have also been used. After considering numerous alternative classification systems for diagnostic information, the Medicare program chose to implement the CMS Hierarchical Condition Category (CMS-HCC) classification system using 70 diagnostic groups for prediction. As of 2011, up to 86 HCCs are used, and the system is also used for Medicare Part D which includes prescription drug plans. Recent research has suggested that more sophisticated risk adjustment has led to less risk selection in the Medicare Advantage market, but insurance companies still have some ability to select low risks. They show that plans can still select on the individuals' risks, *given risk adjustment*. A more complete risk adjustment model that compensates plans for the average risk of as many targetable groups as possible might mitigate this problem.

In the United Kingdom (UK), risk adjustment has been used for many years to allocate funds between geographically defined "Primary Care Trusts" using need, utilization and health status variables, and done at the group level. More recent efforts in the UK have considered using individual information for risk adjusting payments not only to the geographically defined PCTs, but also to individual general practitioners. The main difficulty of using individual-level diagnostic information has been the process of obtaining this information from office based physicians who are not required to record diagnoses as a

condition of service payments, leading to exploration of models that use only inpatient diagnoses and counts of office and facility visits.

Risk adjustment models using a variety of adjusters are also used in all European countries with multiple, competing health plans, as well as in Chile and Colombia.

Econometric Issues

Risk adjustment models have been an active area for testing and developing new econometric methods. Early models used primarily linear models in part because the very large sample sizes and large number of explanatory variables made estimation of nonlinear models time-consuming if not infeasible. Since the 1990s and 2000s, there has been a surge of interest in building robust nonlinear models that are less sensitive to the outliers that are common in highly skewed expenditure data. The two-part log linear model used so widely in the Rand Health Insurance Experiment has been largely laid to rest by several studies that demonstrated the severe problems caused by uncorrected heterogeneity in such models. Among nonlinear models, Cox Proportional Hazard models and Generalized Linear Models are the most widely used.

A central finding in the recent literature is that while nonlinear risk adjustment models may be superior for hypothesis testing, by creating test statistics for hypothesis tests that have well behaved properties, the nonlinear models generally do worse than simple least squares models when used to predict sample and subsample means. Prediction of dependent variable means in levels in nonlinear models is seriously confounded by heteroskedasticity, which can be so multidimensional that it is very difficult to correct in medium sized samples, and estimation of rich nonlinear models in mega samples needed to capture all of this heteroskedasticity are still hampered by the complexity of estimating precise models with hundreds (or even thousands) of parameters on multiple millions of observations. In sum, while non-linear models can potentially produce better estimates, this improvement only comes through making various parametric assumptions. When these assumptions are not satisfied (as is probably often the case since they cannot be tested), simple least squares estimates are better because they do not require any parametric assumptions in order to be unbiased.

In recent years there has been a return of support for least squares models, as signaled by their use by researchers in Australia, Germany, the UK, and the US as well as for practical implementation in Belgium, Israel, Netherlands, and Switzerland. The preferred approach since 2000 for the US Medicare program has consistently been to use weighted least squares regressions of annualized spending on the risk adjusters, where annualized spending is actual spending divided by the fraction of the year a person is

eligible, and this annualized amount is weighted by the fraction of the year a person is eligible to generate unbiased means. Such an approach replicates the mean exactly in disjoint groups, and is the only demonstrated approach that easily accommodates individuals with partial year eligibility. The mega-samples of multiple millions of observations, used to develop Figure 2 and Table 2 in this article, largely alleviate concerns about overfitting of outliers even with great skewness.

Future Directions in Risk Adjustment

Risk adjustment figures prominently in the US Affordable Care Act of 2010, notably in the proposals for establishing insurance exchanges to serve the individual and small group insurance markets. To keep insurance affordable, premium subsidies will be offered by the government, and premium rate bands will limit premium variations across age and gender groups to be no more than three to one. It is readily seen from Figure 1 above that in the absence of regulation, plans would choose to charge 64 year old males a premium that is nearly ten times that of a 10 year old male. Such regulated premiums can only be feasible if premium subsidies to plans are risk adjusted so that plans are paid for enrolling the aged and relatively unhealthy.

The US Department of Health and Human Services has developed new risk adjustment models for use in the insurance exchanges. The biggest changes between the new models and the CMS-HCC model are separate models for adults, children, and infants and the use of a concurrent rather than prospective framework. There are also different weights depending on the “metal level” of the plan due to differing plan liability, which labels plans according to whether they are expected to cover 90 % or more of health care spending (platinum), 80% or more (gold), 70% or more (silver) or less than 70% (bronze). Because plan differences are explicit in terms of what is covered, selection incentives are strong, and aggressive risk adjustment is needed, which may explain the use of a concurrent rather than a prospective framework. The proposed concurrent models have R-squareds from 0.289-0.360, much larger than the typical prospective model which typically has an R-squared around 0.12. The concurrent framework also accommodates the lack of previous diagnostic data for new enrollees. While a concurrent framework solves this problem and substantially improves prediction, it also represents a payment system that begins to look similar to cost-based reimbursement. It will be interesting to see if it will bring some of the moral hazard problems (i.e. upcoding and increased utilization) that come with that type of payment scheme.

The exchanges have also introduced new questions about how risk adjustment interacts with other forms of plan payment. In the exchanges, plans will be paid through age-rated premiums and through risk adjustment, reinsurance, and risk corridor transfers, so the revenues a plan receives for an individual will be the sum of these payments, not just the risk adjustment payment. In the exchanges premiums can vary

by age, and this age-based premium variation can lead to improvements in welfare by causing more efficient sorting of individuals to health plans. However, as risk adjustment and reinsurance payments compensate plans for age-based differences in cost, competition will cause the plans to vary premiums less and less. In the extreme (an extreme which is easily achievable) risk adjustment and reinsurance will fully compensate plans for age-based cost differences and premiums will not vary by age. This will lead to inefficient sorting. Hence, it is clear that when premium-rating is allowed, there is a tradeoff between minimizing selection incentives by maximizing the fit of a payment system and inefficient sorting caused by a lack of variation in premiums. It also remains to be seen how risk adjustment, premiums, and reinsurance will interact with taste heterogeneity.

New areas that are also receiving a great deal of attention are customized risk adjustment models that predict outcomes other than total spending. Predicting hospitalizations, length of stay, hospital resource use, readmissions, mortality, performance measures and primary care service needs are all examples of specific risk adjustment models that have been calibrated and are increasingly being used. Value-Based Payment is another example of a US reform that benefits from risk adjustment. With the new emphasis on detecting and rewarding good performance, risk adjustment is destined to see further expansion in use for these new outcomes globally.

Another important current area for risk adjustment in the United States is in bundled payments to Accountable Care Organizations (ACOs), which are moderate-size health care provider networks willing to receive a bundled payment in exchange for taking responsibility for providing all care to a panel of patients. Given the modest size of these panels, risk adjustment will be critical for ensuring that both healthy and sick enrollees are welcomed in the ACO. Comparing actual to risk-adjusted predictions of various performance outcomes within the ACO is a also key concept in these organizations.

A final important area for risk adjustment is in bundled payment for primary care, particularly as part of the Patient-Centered Medical Home. In this CMS initiative, the Medicare program is encouraging primary care providers to take responsibility for providing comprehensive primary care for patients from all payers (Medicare, Medicaid, and private) and offering increased primary care “base payments” for the extra effort this will take (beyond what they will be reimbursed for via FFS). These base payments will be partial capitation amounts, not fee-based. Sizable bonus payments are also being considered to reward primary care practices for achieving specified quality, cost, and patient satisfaction targets. If either the base payments or bonus payments are not risk adjusted, then primary care practices could potentially act like insurance companies, striving to attract the healthy and avoid the relatively sick, undermining the potential of the PCMH initiative.

To date, risk adjustment models in the US have relied primarily on demographic and claims-based (usually diagnostic) information to adjust payments, utilization and outcome measures. Occasionally self-reported information is used, although the relatively high cost of surveys and consumer input limit the widespread use of such information. A potentially huge source of information for the future are electronic health records, which capture not only what treatments are done, but also the results of various biometric and laboratory tests and imaging procedures. Health records will be challenging to use, but offer rich possibilities for improved prediction of diverse outcomes of key interest to researchers and policymakers.

Glossary

Risk adjustment: the use of information to explain variation in health care spending or other outcomes such as resource utilization, mortality or health over a fixed interval of time, such as a quarter or year

Risk selection: when an individual's choice of insurance or a service is correlated with her cost (risk) to the insurer

Base period: the period from which information is used to predict costs or other outcomes

Prediction period: the period for which the regulator would like to predict an outcome

Concurrent risk adjustment: the use of variables measured in the prediction period to predict outcomes in the same period

Prospective risk adjustment: the use of variables measured in a prior base period to predict outcomes in the prediction period

Optimal risk adjustment: models of risk adjustment that incorporate behavioral objectives in setting plan premiums or other outcome targets, potentially allowing biased predictions

Conventional risk adjustment: risk adjustment models that focus solely on unbiasedness and statistical properties such as maximizing the models explanatory power (R-square)

Service level risk selection: the act of distorting the level of services offered in an insurance contract in order to attract low risks; for example, the exclusion of diabetes specialists in an insurer's network to dissuade high-risk diabetics from enrolling in the plan

Community rating: when insurance premiums are not allowed to vary across individuals

Recommendations for further reading

1. Ash, A., and Ellis, R.P. (2012). Risk-adjusted payment and performance assessment for primary care. *Medical Care* 50, 643-53.
2. Ash, A.S., Porell, F., Gruenberg, L., et al. (1989). Adjusting Medicare capitation payments using prior hospitalization data. *Health Care Financing Review*, 10, 17-29.
3. Breyer, F., Bundorf, M.K., and Pauly, M.V. (2012). Health care spending risk, health insurance, and payments to health plans. In Pauly M.V., McGuire, T.G., and Barros, P.P. (eds.) *Handbook of Health Economics*. Volume II. Elsevier.
4. Bundorf, K., J. Levin, and N. Mahoney. (2012). Pricing and welfare in health plan choice. *American Economic Review*, forthcoming.

5. Cutler, David M. and Richard J. Zeckhauser. (2000). The anatomy of health insurance. In Cuyler, A.J. and Newhouse, J.P. (eds.) *Handbook of Health Economics*. Volume I. Elsevier. 563-637.
6. Dixon, P, Dushieko M, Gravelle H, Martin S, Rice N, et al. (2011). Developing a person-based resource allocation formula for allocations to general practices in England. Nuffield Trust. Downloaded on July 15, 2011 from http://www.nuffieldtrust.org.uk/sites/files/nuffield/document/Developing_a_person-based_resource_allocation_formula_REPORT.pdf
7. Einav, L. and Finkelstein, A. (2011). Selection in insurance markets: Theory and empirics in pictures. *Journal of Economic Perspectives*, 25,115-138.
8. Einav, L., Finkelstein, A., Ryan, S., Schrimpf, P., and Cullen, M.. (2013). Selection on moral hazard in health insurance. *American Economic Review*, 103, 178-219.
9. Ellis, R.P. (2008). Risk adjustment in health care markets: concepts and applications. In Lu, M., and Jonnson, E. (eds.) *Paying for Health Care: New Ideas for a Changing Society*. Wiley-VCH publishers Weinheim , Germany.
10. Ellis, R.P. and McGuire T.G. (2007). Predictability and predictiveness in health care spending. *Journal of Health Economics*, 26, 25-48.
11. Glazer, J., McGuire, T.G., (2000). Optimal risk adjustment of health insurance premiums: An application to managed care. *American Economic Review*, 90, 1055–1071.
12. Iezzoni, L.I., ed. (2013). *Risk Adjustment for Measuring Healthcare Outcomes*, Fourth Edition. Ann Arbor, Michigan: Health Administration Press.
13. Newhouse, J. P. et al. (1989). Adjusting capitation rates using objective health measures and prior utilization. *Health Care Financing Review*, 10, 41-54.
14. Pope, G.C., Kautter, J., Ellis, R.P., Ash, A.S., Ayanian, J.Z., Iezzoni, L.I., Ingber, M.J., Levy, J.M., Robst, J.. (2004). Risk adjustment of Medicare capitation payments using the CMS-HCC model. *Health Care Financing Review*, 25, 119-141.
15. Rothschild, M. and Stiglitz, J. (1976). Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *Quarterly Journal of Economics*, 90, 629-649.
16. Rice, N. and Smith, P. (2001). Capitation and risk adjustment in health care financing: An international progress report. *Milbank Quarterly*, 79, 81-113.
17. Thomas, J.W., Grazier, K.L. and Ward, K. (2004). Comparing accuracy of risk-adjustment methodologies used in economic profiling of physicians, *Inquiry*, 41, 218-231.
18. Van de Ven W.P.M.M., Beck, K., Buchner, F., Chernichovsky, D., Gardiol, L., et al. (2003). Risk adjustment and risk selection on the sickness fund insurance market in five European countries. *Health Policy*, 65, 75-98.

19. Van de Ven, W.P.M.M., Beck, K., Van de Voorde, C., Wasem, J., and Zmora, I. (2007). Risk adjustment and risk selection in Europe: 6 years later. *Health Policy*, 83, 162–179.
20. Van de Ven, W.P.M.M. and 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.
21. Winkelman, R., and Mehmud, S. (2007). *A comparative analysis of claims-based tools for health risk assessment*. Schaumburg, Ill.: Society of Actuaries.

Table 1 Risk Adjustment Models Used for US Public Programs

Model Feature	Adjusted Clinical Groups (ACGs)	Chronic-Illness Disability Payment System (CDPS)	Clinical Risk Groups (CRGs)	Diagnostic Cost Groups (DCG)/ Hierarchical Condition Categories (HCC)	Episode Risk Groups (ERGs)
Background					
Model Developer	Johns Hopkins	University of California, San Diego (UCSD)	3M Health Information Systems	Verisk Health (formerly DxCG)	Ingenix (formerly Symmetry)
Marketplace Introduction	1992	1996	2000	1996	2001
Disease Classification					
Additive/Categorical Classification	Categorical	Additive	Categorical	Additive	Additive
Users:					
Government Programs	4 Medicaid	10 Medicaid	1 Medicaid	Medicare 1 Medicaid	1 Medicaid
Commercial (in 2009)	175	None	7	300+	60
Prospective R-Squares:					
Without Truncation	16.60%	14.70%	N/A	17.80%	16.40%
Truncated at \$100,000	21.80%	20.80%	N/A	24.90%	24.40%

Notes: For further details of each of the systems listed, see 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, 453-472; Kronick RT, Dreyfus, T, Zhou Z. 1996. Diagnostic risk adjustment for Medicaid: the disability payment system. *Health Care Fin Rev* 17, 7-33; Averill RF, Goldfield NI, Eisenhandler J, et al. 1999. Development and evaluation of clinical risk groups (CRGs). Wallingford, CT: 3M Health Information Systems.; Ash, A.S., Ellis, R.P., Pope, et al. 2000. Using diagnoses to describe populations and predict costs. *Health Care Fin Rev*, Spring 21(3): 7-28.; and Symmetry Health Data Systems, Inc. 2001. Episode risk groups: ERG user’s guide. Phoenix, AZ: Symmetry Health Data Systems, Inc. Prospective R-Squares are from Winkelman and Mehmud (2007). Characterization of each system is from Florida Agency for Health Care Administration. 2009. “Risk Adjustment Model Comparison” Downloaded on October 17, 2011 from http://ahca.myflorida.com/Medicaid/quality_management/workgroups/managed_care/5_rar_model_comparison_050709.pdf.

Table 2. Results from the United Kingdom Predicting FY2008 Health Spending per Capita Using Prior Two Years of Data

ID	<u>Explanatory variables in OLS models:</u>	<u>Number of parameters</u>	<u>Individual level R²</u>		<u>Practice level R²</u>
			Estimation Sample <u>N=5,206,651</u>	Validation Sample #1 <u>N=5,205,747</u>	Validation Sample #2 <u>N=797</u>
a.	Age and gender only	38	0.0373	0.0366	0.3444
b.	Model (a) plus 152 diagnosis groups and 4 lagged utilization variables	194	0.2656	0.2610	0.7394
c.	Model (b) plus 151 geographic dummies	345	0.2659	0.2612	0.8046
d.	Model (c) plus 135 attributed need and 63 supply variables	543	0.2662	0.2615	0.8254
e.	Model (c) plus 7 attributed need and 3 supply variables	355	0.2671	0.2622	0.8254
f.	Age/gender, 152 diagnosis groups, 151 geographic dummies, 7 attributed need and 3 supply variables	351	0.1272	0.1229	0.7738

Notes: Diagnosis groups use only inpatient diagnoses from a two prior years. Utilization variables include inpatient episode count, outpatient visit count, dummy=1 if any priority referral, and dummy=1 if any outpatient visit; all measures are for prior two years. Estimation sample is a 10% random sample of the UK population. Validation Sample #1 is a different 10% random sample of the UK population drawn without replacement. Validation Sample #2 is a 100% sample of patients at 10% of primary care practices. All results are from Dixon et al (2011), especially Table 7.4 and Appendix 13, Table 9. http://www.nuffieldtrust.org.uk/sites/files/nuffield/document/Developing_a_person-based_resource_allocation_formula_REPORT.pdf

Figure 1. Indifference curves between consumption in two states of the world using the Rothschild and Stiglitz (1976) framework, as in Cutler and Zeckhauser (2000).

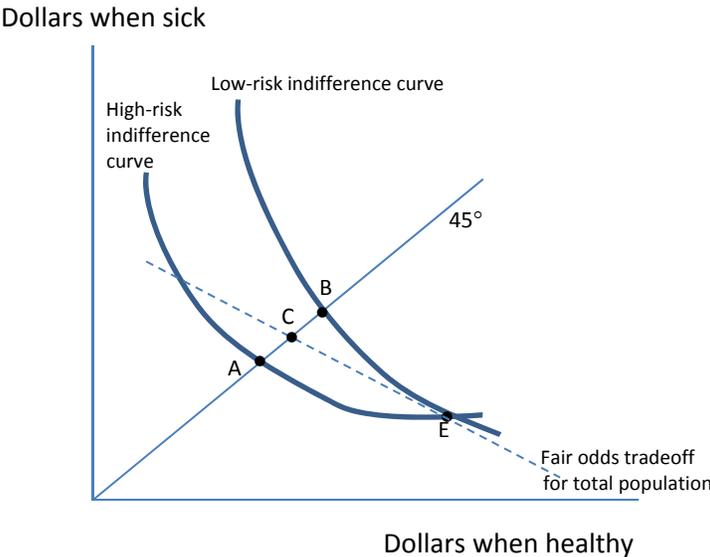
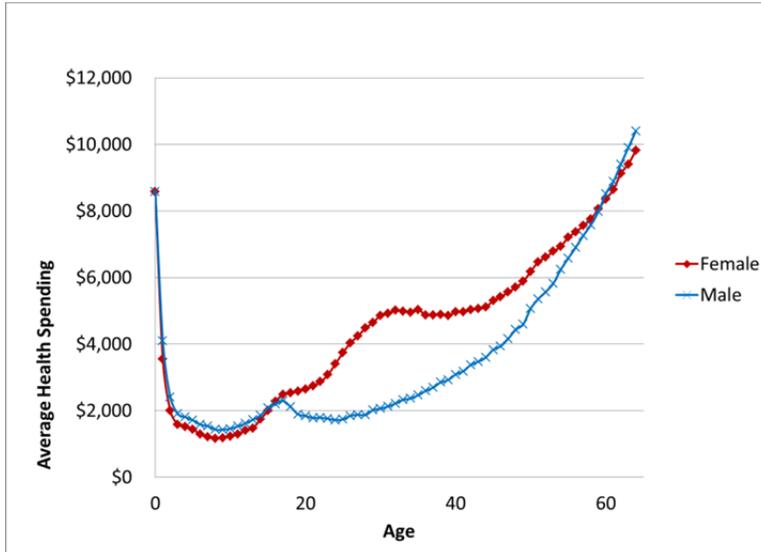


Figure 2: Actual Spending by Age and Gender in US, 2009



Notes: Sample used is the US 2009 Truven MarketScan commercially insured claims and encounter data. All plan types and individual with a valid sex and age <65 were included, although persons without pharmacy coverage were excluded. Each point plotted is the one year average total covered health spending per capita (medical plus pharmacy spending, including deductibles and copayments, but excluding dental and vision spending) for that one year age and gender group. (Source: Authors' original figure.)