CHAPTER ELEVEN

Health Care Spending Risk, Health Insurance, and Payment to Health Plans

Friedrich Breyer, M. Kate Bundorf and Mark V. Pauly

Contents

Handbook of Health Economics, Volume 2
ISSN: 1574-0064, DOI: 10.1016/B978-0-444-53592-4.00011-6

Abstract

This chapter will deal with the actual and efficient functioning of health insurance in settings where risk (expected value) of medical spending or insurance benefits varies across individuals at a given point in time or over time for a given individual. It will deal with equilibrium in insurance markets with risk variation and will also deal with various configurations of information, the impacts on such markets of regulation motivated by risk variation, and the actual and optimal impact of governmental policies to deal with risk variation in national insurance systems.

Keywords: health care spending; health care markets; insurance; public policy; risk variation

JEL Code: I1

1. VARIATION IN HEALTH CARE SPENDING BETWEEN PERSONS AND OVER TIME

Annual spending on health care varies significantly across individuals. The top 5 percent of spenders account for approximately half of US health care

expenditures in a given year (Zuvekas and Cohen, 2007). That statistic is, however, misleading as a measure of the extent of risk variation because risk, when defined as expected value, represents the predictable component of spending or claims, and not all or even most of actual spending variation across people or over time is predictable. Table 11.1, taken from Pauly and Herring (1999), compares actual and predicted variation in spending on insured medical expenses for non-elderly individuals without public insurance in the US in 1987. There is substantial variation in both actual spending and in risk, but the two are not the same, with much less of the distribution of risk concentrated on people with close to zero or very high actual spending.

The extent to which and for whom risk variation is predictable has important implications for health insurance markets. While the unpredictable component of variation in expenditures generates demand for insurance, and, indeed, is the reason for

Table 11.1 Key Statistics on Actual and Predicted Insurable Expenses*

A: Key Statistics on Actual Insurable Expenses	
Coefficient of variation = 377.80	
Mean	$797
99th percentile	$12,811
95th percentile	$3,220
90th percentile	$1,549
75th percentile	$450
50th percentile	$137
25th percentile	$22
10th percentile	$0
B: Key Statistics on Predicted Individual Insurable Expenses: Linear Prediction Model, Using 1986 Conditions	
Coefficient of variation = 85.61	
Mean	$822
99th percentile	$3,710
95th percentile	$2,263
90th percentile	$1,592
75th percentile	$991
50th percentile	$647
25th percentile	$402
10th percentile	$223

Note: Number of observations = 8,010. All amounts are in 1987 dollars. Only insured individuals under age 65 with no public assistance are included.

*Pauly and Herring (1999), p. 31 (Table 3-1).

Source: NMES data.

insurance to exist, the predictable component influences the functioning of insurance markets. At one extreme, if almost none of the variation is predictable by anyone, premiums will be uniform for any nominal level of coverage and competitive markets with voluntary insurance purchase will work well, so that those persons unlucky enough to have high spending will have their financial risk cushioned and their consumption smoothed. At the other extreme, if each person's spending is exactly predictable by everyone, there would be no utility gains from voluntary insurance but there may be policy pressure to improve equity by redistributing premium costs from low risks to high risks.

When the power to predict risk is between these two extremes, there are two concerns that motivate potential public intervention to improve welfare. The standard concern in theoretical models of competitive insurance markets of all types is that, if consumers have and use more information about their expected loss than insurers do, adverse selection may limit efficiency. In markets where insurers can predict risk at least as well as consumers can, there are policy concerns about fairness in health insurance premiums, and about mitigation of lifetime risk of the variation of premiums imposed in single period settings, the so-called "reclassification risk" (Hirshleifer, 1971). Here the issue is whether unregulated insurance markets develop mechanisms to protect against such risk, and there is considerable evidence that they do.

Regulators need as much risk information as insurers have to play a proper role both in setting payments to plans and in regulating enrollee premiums. If insurers know more than regulators but premiums are not varied with risk, insurers may engage in cream skimming. Thus, the relative predictability of health expenditures from the perspectives of insurers, regulators, and consumers is a potentially important determinant of the functioning of health insurance markets and of health insurance regulation.

While risk variation thus is the motivation for a great deal of research and policy advice in this area, it is important to note at the outset that the contexts for this research and advice differ widely. Both in terms of global prevalence and certainly in terms of research weight, the empirical evidence on the effects of risk variation in insurance markets is nearly always based on a situation in which regulation or some other policy is already present that limits or forbids insurers from tailoring current-period premiums paid by consumers specifically to risk. Often as well there are substantial subsidies to the insurance purchased by almost everyone in the market. Other work, almost all conceptual and much smaller in quantity, deals with potential inefficiencies in private health insurance markets without regulation or other policy constraints, compared to what public policy could or should do as a way of improving matters. In what follows we will attempt to be clear which situation is furnishing the context for the research being summarized and discussed.

1.1. Cross-sectional Variation across Individuals

Most empirical studies of the predictability of health expenditures have been conducted from the perspective of regulators who wish to risk adjust payments to health plans whose explicit premiums (if any) are required to be uniform across buyers (community rated). We begin with analysis of US data from the perspective of regulation of its Medicare program. In the early 1980s, the Medicare program for the aged and disabled began offering beneficiaries the opportunity to replace their traditional publicly managed insurance benefits with coverage from a private plan, and the introduction of the program led to one of the first attempts to implement a system of risk adjustment.

In response to the concern that the initial methods used by government regulators to risk adjust payments to private plans were inadequate, early studies sought to calculate the maximum amount of variation in annual health expenditures which was potentially explainable in order to provide a benchmark of how well a system of external risk adjustment could in theory perform (Newhouse et al., 1989; van Vliet, 1992). The perspective is primarily that of a regulator that wishes to adjust payments received by insurers based on the risk of those they insure. Using panel data from the Rand Health Insurance Experiment, Newhouse et al. (1989) decomposed the variation in observed expenditures into the within- and between-person components, and identified the between-person variance as potentially explainable. Recognizing, however, that an enrollee may be able to anticipate at least some changes over time in health care expenditures (a classic example is pregnancy, although it is not very relevant to Medicare), they considered the between-person component to be a lower bound estimate of the maximum variation either insurers or beneficiaries could explain or predict. Studies adopting methods to model autocorrelation in individual spending (including regression to the mean) found that, while such autocorrelation existed, the additional explanatory power of these models was not substantially higher (Welch, 1985; Newhouse et al., 1989; van Vliet, 1992). Taken together, these studies suggested that the maximum explainable between-person variation was between 15 and 20 percent of total health expenditures, and that explainable variation was higher for outpatient care and pharmaceuticals than for inpatient spending. They also demonstrated that the types of risk adjusters initially used by the Medicare program (age, sex, Medicaid status, institutional status, and employment status) explained only about 1 percent of variation in expenditures. Thus, by providing a benchmark against which Medicare's then-limited risk adjustment scheme could be compared, this research demonstrated that regulated payments likely reflected only a subset of the information available to insurers and potentially to regulators.

1.1.1. Observable Individual Characteristics which Explain Expenditure Variation

Studies consistently demonstrate that, while health expenditures are highly correlated with the demographic characteristics of age and sex, demographic variables alone explain a very small portion—about 1 to 5 percent—of variation in health spending across individuals or households (Newhouse et al., 1989; Ellis and McGuire, 2007; Prinsze and van Vliet, 2007; Lamers, 2001). Indicators of chronic conditions (direct measures of the presence of conditions or indicators of prior-period spending on treatments for them) substantially increase the explanatory power of models of health expenditures (Newhouse et al., 1989; Shen and Ellis, 2002; Luft and Dudley, 2003; Ellis and McGuire, 2007), but still leave much unexplained. For example, Ellis and McGuire (2007) estimate that adding diagnostic indicators to a model with initial information only on age and sex increases the *R*-squared from 0.01 to 0.11. However, many studies find that measures of prior use make an independent contribution to the explanatory power of models, even after controlling for the presence of chronic conditions (Newhouse et al., 1989; Lamers, 2001; Shen and Ellis, 2002; Hsu et al., 2009). An exception is Ellis and McGuire (2007), who find that including prior year use has little effect on the explanatory power of their models. Rather they find that models estimated separately by service type, using diagnostic measures, have greater explanatory power than those estimated including prior spending related to total expenditures.

Relatively few studies have examined the correlation of spending with other (non-medical) individual characteristics such as socioeconomic status and medical care preferences, particularly after controlling for measures of health status. This is probably because many studies use insurance claims data which generally include relatively few socioeconomic or preference measures. However, if spending varies positively with income at any level of coinsurance, then expected expenses will be higher for people with higher incomes. If people at different income levels buy the same nominal insurance policy at a uniform premium, higher income people, expecting higher benefit payments, will for this reason alone be more likely to seek insurance coverage. There will be adverse selection if premiums do not vary with income, even though there is no variation in health risk. We will discuss this point further below.

While Newhouse et al. (1989) find that a limited set of subjective (self-assessed) measures of health have relatively little incremental explanatory power after controlling for measures of health status generated from claims data, there are many other characteristics of individuals not so easily observable in claims data that are potentially important determinants of expected expenditures. For example, Cohen et al. (2006) find that individual characteristics such as family income and marital status predict the likelihood of high spending in a given year, even after controlling for prior year health expenditures. Cutler et al. (2008) document a correlation between measures of risk tolerance and medical expenditures and use of medical care. Thus, while these types of individual characteristics are correlated with health expenditures, little evidence

exists on their quantitative importance in predicting expenditures, particularly after controlling for health status.

1.1.2. How Much Variation is Predictable by Consumers?

Relatively little is known about how well consumers are able to predict their future health expenditures. On the one hand, the consumer potentially has the most complete set of information about next-period spending. Not only do consumers know their prior utilization, which the literature indicates is correlated with future expenditures, but they also have private information on how their utilization is likely to change in the next time period. Potential examples include health measures like pregnancy, elective surgery, or genetic risk, and preference measures like whether the person is a hypochondriac or alternatively prefers to avoid hospitals at all costs. On the other hand, consumers may be the least sophisticated users of some types of information. For example, consumers may be uninformed about the relationship between age and expected expenditure and thus may not accurately incorporate aging into their forecast.

What is relevant to the possibility of adverse selection and market equilibrium is not whether consumers can predict their future expenses accurately overall, or more precisely than insurers can, but rather whether they know or know more accurately important indicators of risk that insurers do not, and how they interpret and use this information, even if imperfectly, to judge their likely claims under different insurance policies. Competition among insurers would cause premiums to reflect everything insurers know about risk even if consumers were totally ignorant, but adverse selection could still arise if consumers had better information on other important variables that predict future spending, such as their conditions or preferences. On the other hand, if consumers estimate their risk levels differently than insurers do (perhaps because they do not know how to convert current period characteristics into future expected expenses), they may decline coverage offered at accurately risk-rated premiums and modest administrative expenses, or may seek overly generous coverage at high administrative expense levels. Still, as risk adjustment becomes more sophisticated, it is entirely possible that the regulator or the insurer may know more and more accurate information than any information the consumer might use. For example, an insurer or regulator who has population-level information on the relationship between the presence of chronic disease and health expenditures may more accurately forecast the impact of a newly developed chronic condition on future spending than could the consumer. (Consumers may still know their preferences and symptoms, however.) Yet we are unaware of any direct evidence on what types of information consumers use when forecasting health expenditures. Of course, if consumers believe that insurers always set their premiums close to expected benefits, then those faced

with a high premium will modify any subjective judgment and assume that they must be high risk.

A fundamental question then is whether health insurance buyers in some markets actually do know useful information about risk that insurers do not. If insurers are forbidden to charge premiums based on risk variables they know, there can still be inefficient adverse selection. And the inefficiency associated with mispricing would depend on how responsive demand is to that mispricing. Conversely, for various reasons to be discussed, consumers may sort by risk but this is not necessarily inefficient, especially if there was no information asymmetry that motivated the sorting (Chiappori and Salanié, 2000; Einav et al., 2007). The gold standard would then be direct measurement of relative information about risk predictors, but this is rarely available. Instead, the best that can be done is to look at selection controlling for information supposedly observed by the insurer—to see if there is selection based on information not available to the insurer (Chiappori and Salanié, 2000).

1.1.3. How Much Variation is Predictable by Insurers and Regulators?

At the time of enrollment, insurers and regulators in theory have access to similar types of information about an individual's expected benefits. In practice, when insurers face relatively few regulatory restrictions on the use of information in underwriting, they ask applicants relatively detailed questions about their and their family's medical history and pre-existing medical conditions. In some cases, they have applicants undergo a medical examination or provide a physician contact. For example, an application from a US commercial insurer selling coverage in the individual market asks whether the applicant has received care for or had symptoms of 24 different conditions over the last 20 years. The applicant must provide detailed information on the treatment of those conditions, including the contact information of health care providers. The application also includes questions about tobacco use, alcohol consumption, prescription drug use over the last 12 months, and whether the applicant has ever had any application for either health or life insurance restricted in any way. In many countries, insurers have the right to cancel individually purchased insurance if they find that the applicant provided incomplete or inaccurate information. Requirements for information and aggressiveness of cancellations (rescissions) vary across insurers. Usually, however, insurers who collect less information charge higher premiums.

Little formal evidence exists, however, on how insurers use this information. Insurers may use the information to decide whom to offer coverage, what services to cover, and what premium to charge. Some German health insurers provide lists of percentage surcharges for certain conditions (e.g. gallstones 40 percent, gout 40 percent, cervical spine syndrome 40 percent, thyroid hyperfunction 30 percent, hay fever 20 percent) and for denial of coverage (e.g. epilepsy, mental illness, Crohn's disease, cancer within seven years). However, because underwriting practices are an integral part of an

insurer's business strategy, most treat them as a confidential business practice, and whether and how insurers use different types of information likely varies across carriers. Anecdotal information from insurers suggests that many companies use the information they collect from applicants to develop a risk adjustment factor, which indicates an applicant's expected expenses relative to the average premium. For applicants with a risk adjustment factor above a particular threshold (in Germany, 500 percent of baseline), an insurer is likely to decline making an offer of coverage. For others, insurers may customize the terms of coverage. Some insiders indicate that, in the absence of strict regulations governing how they use information, insurers use the detailed information they gather in a relatively nuanced way. For example, an applicant whose diabetes is well controlled would be treated differently from one whose is not. Thus, underwriting in the US is based in part on objective data but in part on the judgment the underwriter makes after viewing all available information; premium adjustments are usually tied in a predetermined way to objective data (like the presence of a chronic condition) but an underwriter can turn down an application entirely based on judgment.

In the US market, the level of detail in the information insurers collect and use when setting coverage terms varies based on the size of the group. While small group insurance was risk rated at the level of the group before regulations in most states limited it, there was less attention to risk than in the case of individual insurance. In individual insurance markets without regulation, after adjusting for age and sex, insurers will typically rate new applicants as "standard" or "non-standard" based on health information, and increase premiums for the latter group by up to 200 percent of standard before declining to insure entirely (Pauly and Herring, 1999).

Another possibility is that insurers or insureds may have information on genetic risk from genetic tests. Of course, insurers have traditionally collected as part of the underwriting process information that included the answers to questions about causes of death for parents and siblings, which of course is genetic information. But the advent of genetic tests has raised questions about substantial improvements in the ability to predict future medical costs. Up to this point, however, genetic tests have not been a major factor in underwriting, largely because they are not much better than is family history at predicting costs in the next time period (but before the person goes on Medicare), which is what the insurer wants to know. At best they predict what illnesses the person eventually may have, but not precisely when. There are some exceptions for tests for relatively rare breast cancer and cystic fibrosis, but in general insurers professed to be not interested in using such tests themselves—although paradoxically they were concerned when the insurance purchaser had test results (Subramanian et al., 1999). Nevertheless, laws often now do greatly limit insurers' ability to ask for and use such information.

Subsequent to enrollment, however, an insurer can accumulate extensive information regarding an individual's risk because the insurer observes the person's claims.

Studies suggest that, subsequent to enrollment, the risk adjustment systems used by regulators represent a subset of the information available to insurers. Shen and Ellis (2002) simulate potential insurer profitability associated with effective insurer selection under differing assumptions about the information sets available to insurers and regulators. They find that even when regulators use fairly sophisticated risk adjustment, insurers can still identify individuals who can generate costs in excess of the risk-adjusted payments. Although this study does not provide evidence on what plans actually do, the results indicate that they face relatively strong incentives to seek to induce some individuals to disenroll. In contrast, as will be discussed in more detail below, the bulk of individual insurance in the United States is written under a contract provision in which the insurer promises not to use any such information to single out insureds for premium changes or refusals to renew coverage.

1.2. Persistence in Health Expenditures

Estimates of the maximum explainable variation are based in part on the degree of persistence (serial correlation) in individual spending. Both van Vliet (1992) and Newhouse et al. (1989) present evidence that expenditures are somewhat, but not particularly highly, correlated over time. For example, van Vliet (1992) calculated a correlation coefficient of approximately 0.25 for year one and year two spending and approximately 0.10 for year one and years three through five spending. While subsequent literature uses different methods to quantify the degree of persistence, they confirm this general finding of some, but limited correlation in, year-to-year expenditures. For example, using claims data from a sample of large, self-insured firms, Eichner et al. (1997) find that average expenditures for enrollees in the top decile of spending in a given year were $11,249, or eight times the sample average. Expenditures of the highest spending group in the initial year remained higher than, but moved substantially closer to, the average in subsequent years. In the following two years, they were approximately 5 and 3 times average spending. Similarly, average expenditures for those with the lowest expenditures in the initial year—zero spending for the three lowest deciles—increased to about the sample average in subsequent years. Using two years of data from the nationally representative US Medical Expenditure Panel Survey, Monheit (2003) finds that, of the top 5 percent of spenders in 1996, 30 percent were in the top 5 percent of spenders in 1997. These studies also find that the persistence of health care spending increases with age (Eichner et al., 1997; Monheit, 2003; Pauly and Zeng, 2004), but that the degree of persistence among older adults is limited by mortality, which generates a sharp increase in spending in the last year of life (Garber et al., 1998). Spending persistence also varies significantly across types of services. For example, spending on prescription drugs is much more persistent than total spending or hospital spending (Pauly and Zeng, 2004).

To date, empirical analyses of persistence are based on longitudinal data spanning time periods from two to eight years. Yet evidence from a longer time period, even a lifetime perspective, is important for evaluating the efficiency and distributional implications of different forms of health insurance. Using data from a three-year panel to simulate health expenditures over a working lifetime, Eichner et al. (1997) find that concentration declines significantly over time. Based on their simulations, the percentage of the population accounting for 80 percent of spending increases from 10 to 29 percent and to 48 percent over one, five, and 35 years, respectively.

1.3. Implications

Overall, these studies of predictability and persistence in health care spending indicate that a significant portion of the cross-sectional variation in annual health care spending is unpredictable to consumers and insurers, suggesting that insurance does provide an important benefit in the form of a reduction in the uncertainty of spending on medical care. (In the case of independent events, insurers can profit without knowing which individuals will turn out to be high cost by charging a premium that more than covers average or expected expense.) While the degree of unexplained variation supports the desirability of insurance, evidence of the existence of differences across consumers in the extent to which spending is predictable suggests that, in theory, asymmetric information could threaten the functioning of competitive insurance markets. In addition, because some people are unlucky enough to contract chronic conditions which persist over years and are associated with high spending, they could suffer from high premiums for a long period of time. In practice, however, because studies provide little information on the extent to which information is asymmetric between health insurance consumers and insurers, this literature offers little insight on the degree to which adverse selection is likely to occur. An exception is the above-mentioned study by Pauly and Zeng (2004), which demonstrates that, because spending on prescription drugs is highly persistent, unsubsidized stand-alone insurance is unlikely to be feasible due to adverse selection. Either bundling prescription drug coverage with insurance for other types of medical care or high subsidies are necessary to provide such coverage. This literature provides stronger evidence of a potential gap between insurers and regulators in the predictability of spending. When neither insurers nor regulators have information on prior spending, they in theory have access to similar information when an individual initially enrolls. Subsequent to enrollment, however, the literature identifying the importance of prior spending in explaining future spending suggests that an individual's claims history will provide an insurer with more information on likely expenditures than has been incorporated in most risk adjustment systems. While one way to address this would be to introduce measures of prior utilization into a system of risk adjustment,

a problem is that some predictors of higher spending may themselves be manipulated by beneficiaries or insurers. For example, to the extent that an insurer can gain from increasing prior year spending or utilization that is used in a risk adjustment system, that system will cause inefficiency in health care delivery as it attempts to reduce inefficiency due to risk selection. (The ultimate cause of the risk selection was a decision to limit the extent to which premiums vary with risk.) Thus, the extent to which systems of risk adjustment can be refined to address this gap is an important area of research (see section 4.3.5 below).

It also important to acknowledge that estimates of both maximum explainable variation and spending persistence are specific to both the population and the time period studied. The degree to which health expenditures are predictable likely depends both on the prevalence of chronic relative to acute disease within a population and on the availability and use of technology. For example, in the context of the US Medicaid program, Kronick and Dreyfus (1996) demonstrate that health care expenditures are more predictable for a disabled than for an average population. The reduction in relative prevalence of infectious compared to chronic disease on a population basis has likely made medical care expenditures more predictable—by regulators, insurers, and consumers—and the greater prevalence of chronic diseases in older populations likewise potentially makes their spending more predictable. The future effects of technological change on predictability and persistence, however, are more difficult to predict a priori because different types of technological change would have different implications for expenditure persistence. For example, the development of treatments which reduce mortality from chronic conditions without affecting symptoms can make health care spending to alleviate those symptoms more persistent, while those which prevent or completely cure formerly chronic conditions will make spending less persistent. And technological change may even have different impacts on spending persistence and predictability. Empirical evidence is lacking, however, on the extent to which the predictability of health expenditures has changed over time, the causes of any changes or the implications for the extent and form of insurance coverage.

2. INSURANCE MARKET EQUILIBRIUM WITH RISK VARIATION AND RATING VARIATION: MARKET EQUILIBRIUM BENCHMARKS

Health insurance markets throughout the world are heavily regulated. In order to understand the effects of regulation, it is useful to consider the market equilibrium in both unregulated, competitive markets and regulated markets as benchmarks. The

model which is typically used to study the insurance market is based on the following assumptions (see, e.g., Dionne and Doherty, 1992):

1. The risk of illness can be described by a probability distribution of the possible health state-associated loss for each individual. In the simplest case, there are only two future states of the world, and the individual can become sick and will then incur a purely financial loss (cost of medical care) of M, with probability π, otherwise he or she stays healthy with health care costs of zero.
2. Individuals can differ in their risk of illness π (but not in the size of the possible loss M). In the simplest case, there are only two risk types, $i = H, L$, with $\pi_H > \pi_L$.
3. The individual is risk-averse and will always buy full coverage if this is available at the fair premium, $\pi_i \cdot M$.
4. Initially individuals do not differ in risk aversion, demand for medical care conditional on illness, or preferences for other aspects of insurance (such as extent and form of managed care limitations, provision of information about disease management, or rapid and convenient claims payment).
5. There are a large number of insurance companies who can select what policy or coverage to offer but which, for a given level of coverage, then charge premiums which yield zero economic profits.

We first characterize the market equilibrium within this simple framework to establish a benchmark. We then consider the implications of a series of deviations from this simple model including the effects of costly information on risk, consumer heterogeneity in characteristics other than risk that influence demand for coverage, and asymmetric information between insurers and consumers.

2.1. Equilibrium with Single-period Insurance and with No Regulation and No Information Asymmetry: Underwriting and Risk Rating

2.1.1. Costless Risk Information

In the simplest case, the insurer can observe the probability of illness π_i over a particular time period for each individual i and can therefore calculate the expected loss (or fair premium level) $\pi_i \cdot M$. In the absence of administrative costs (i.e. costs of the underwriting process, selling costs and commissions, costs of billing and of claims processing), competition will drive down premiums for prospective new buyers in that time period towards the fair premium, and the consumer will thus buy full coverage. If buyers differ in risk and insurers can costlessly determine each buyer's risk level, premiums will vary in proportion to risk. For new buyers of single-period insurance coverage, a competitive equilibrium will exist and will be Pareto optimal.

2.1.2. Costly Information

Acquisition and processing of information is costly, and thus interesting questions are how much risk discrimination is efficient and how much is chosen by insurance

companies that need to cover these costs. To answer the first question, it is useful to consider two extremes. If insurance companies do not try to acquire any information on the different risk levels, but applicants know their individual π_i values, then we are in a typical "asymmetric information" situation leading to adverse selection, which will be analyzed below (in section 2.2.1) and in which the efficiency or even existence of market equilibrium is doubtful. If, on the other hand, insurers invest as many resources as needed to perfectly measure the individual π_is (given the information available to the consumers), then this likely results in an inefficient allocation as well since the last unit of information costs primarily serves to discriminate risks, may have little marginal value in doing that, but has even less value in inhibiting adverse selection. Thus there must be some interior private and social optima (not necessarily the same) for the amount of information acquired by the insurers. That is, insurers will not bother to seek information on risk if the risk variation is small or if the information is costly—but they may carry the search for information further than is socially optimal. That is, while they clearly have an incentive to seek information that prevents inefficient adverse selection, they may seek costly information even if adverse selection might be small (for example, because consumers are not aware of risk differences) (Crocker and Snow, 1985). The inefficiency is the use of resources to engage in risk discrimination, not the risk discrimination *per se*. Paradoxically, if insurance buyers are not very price responsive so there will be little inefficiency from adverse selection, there may be offsetting inefficiency as insurers seek excessive information to assist in cream skimming. If there is some information which buyers cannot or will not get that discriminates among risks, insurers will choose to acquire too much of such information, since such information only serves to redistribute (discriminate) among risks.

Often insurers are limited by law, regulation, or custom in using information that predicts risk that they do have or can obtain at low cost. This will lead to an adverse selection outcome even if there is no information asymmetry. The efficient outcome described in the first case will not occur because of regulation.

2.1.3. Impact of Other Insurance and Medical Care Demand Determinants

Insurance demand is determined by more than just the relationship of premiums to expected benefits. The degree to which these other influences are correlated with risk will influence the correlation between risk and coverage in the market, potentially creating a correlation between risk and coverage even in the absence of asymmetric information (see section 2.2). One possibility is that individuals may vary in risk aversion, and that more risk-averse people may be lower risks. Finkelstein and McGarry (2006) find evidence of this type of relationship in the market for long-term care insurance. Fang et al. (2008) find a similar effect in the market for Medicare supplemental insurance, but they attribute it to differences in cognition rather than to

differences in risk aversion. People with poor cognition have poorer health, but their poor cognition also means that they do not recognize the high net value of community-rated insurance to them, so cognition is positively correlated with demand for coverage but negatively correlated with expected utilization.

2.2. Equilibrium with Information Asymmetry: The Benchmark Adverse Selection Model and its Relevance to Health Insurance

2.2.1. Rothschild–Stiglitz Models with Managed Care and Cost Sharing

Various equilibrium concepts have been proposed for the typical adverse-selection situation in which applicants know their risk type but insurers can observe only the share of high risks in the applicant population, μ, but not the risk type of any given individual. An equilibrium à la Rothschild and Stiglitz (1976) is a set of contracts offered by insurance companies with the following properties:

1. Each contract yields a non-negative expected profit to the insurer;
2. There is no potential contract outside this set which would yield a non-negative expected profit to the insurer;
3. Each individual picks the contract which maximizes his expected utility among those offered.

In such an equilibrium with only two risk levels, two contracts are offered: one with full coverage and premium equal to $\pi_H \cdot M$, which is purchased by all high-risk applicants and thus just breaks even, and one with partial coverage M' ($M' < M$) and premium $\pi_L \cdot M'$, which is bought by all low-risk types and also breaks even. The crucial precondition for the self-selection mechanism of the risk types is that M' must be sufficiently lower than M to make it unattractive to high risks (compared to the full coverage, higher-priced alternative). A problem with this equilibrium is that its existence requires the share of high risks, μ, to be sufficiently large. If this condition is violated, the model is without equilibrium and so is silent about what will actually happen at any point in time.

2.2.2. Wilson–Spence–Miyazaki Equilibrium

The equilibrium concept introduced by Rothschild and Stiglitz (1976) has been criticized on several accounts. First, it requires that insurance companies make non-negative profits with each contract they offer. This assumption is not very realistic because it ignores the ability of firms to cross-subsidize among their contracts. Second, it ignores the ability of firms to anticipate the reaction of their competitors to a market entry. Taking both points into account, Wilson (1977), Spence (1978), and Miyazaki (1977) have proposed an alternative equilibrium concept, which is based upon the following assumptions:

1. Each insurer offers a bundle of contracts which, as a whole, yields a non-negative expected profit;

2. No potential bundle of contracts outside the set defined in (1) would yield a positive expected profit if those bundles of contracts which would become unprofitable in response to this offer were taken from the market;
3. Each individual chooses among all contracts offered the one which maximizes his expected utility.

The crucial point which distinguishes this concept from the one by Rothschild and Stiglitz (1976) is that insurance firms contemplating market entry do anticipate the reaction of the incumbent firms. Furthermore, low risks can cross-subsidize high risks. Both features can be interpreted as steps towards more realism in markets with relatively small numbers of foresighted insurers.

Unlike the Rothschild–Stiglitz equilibrium, a Wilson–Spence–Miyazaki equilibrium always exists. It is also a separating equilibrium in which high risks get full coverage but pay a premium lower than $\pi_H \cdot M$, thus they are subsidized by the low risks. In effect, this payment is a bribe from low risks to high risks to keep the latter from further distorting the market for low risks.

Thus the implications of the two equilibrium concepts differ sharply. Unfortunately, the question of whether any of these models describe well the market for health insurance, even an unregulated individual insurance market, is hard to resolve. The possibility of cross-subsidization seems plausible as insurers can use this as a means to increase expected profits. However, the assumption that insurers anticipate the withdrawal of competing contracts in response to their own contract offer remains controversial. Hellwig (1987) has analyzed this issue in a game-theoretic framework which explicitly allows for withdrawals. He finds that the order in which firms move is crucial. However, it is not clear that such an order can be observed in the market for health insurance; indeed, it may not exist at all. Newhouse (1996, pp. 1242ff.) invokes positive costs of writing separate contracts with the low-risk group to generate a pooling equilibrium in which the low risks get their optimal coverage and the high risks get their best feasible coverage.

2.2.3. Equilibrium with Managed Care Plans

The ability of managed care plans to adjust dimensions of insurance other than coverage of a predetermined loss means that there can be a different kind of equilibrium in health insurance. Glazer and McGuire (2000) extend the Rothschild and Stiglitz model to a managed care setting by modeling a health plan as a contract which offers a predefined level of two different types of care, acute and chronic, contingent upon the health status of an enrollee. Consumers are defined by their probability of contracting either an acute or chronic illness, and high risks are those with a higher probability of contracting the chronic condition. In their model, a separating equilibrium exists under the usual conditions with managed care plans underproviding chronic care to low risks. Chernew and Frick (1999) consider the effect of managed care on

the existence and form of equilibria by introducing contracts characterized by two attributes, patient cost sharing and the degree of "managedness," into the R&S framework. "Managedness" refers to a set of non-financial (with respect to patients) restrictions on utilization which could take the form of, for example, utilization review, constraints on physician choice, and provider payment methods. They find that, while the introduction of this second dimension improves the ability of low risks to distinguish themselves from high risks, this ability has ambiguous effects on the existence of equilibria. In some cases, an equilibrium exists when it would not when insurers were limited to a single mechanism. In other cases, it allows an insurer to generate a pooling policy which could break a separating equilibrium, leading to non-existence of equilibrium. Thus, the ultimate effect of managed care on the existence of equilibria in unregulated insurance markets is indeterminate.

2.2.4. Asymmetric Information on Total Coverage

In both equilibrium types described above, a differentiation of the premium per unit of coverage is possible because the insurers are assumed to observe total coverage of each applicant. In other words, they have the power to "ration" health insurance by offering "price-quantity contracts." As a consequence, the price schedule is not linear, but convex: greater coverage implies a higher per-unit price. If total coverage were unobservable, high-risk applicants desiring to fully insure could buy two or more contracts with partial coverage from different insurance companies and thus avoid the steep part of the price schedule. Hence, only linear price schedules would be enforceable, which would render the equilibria described above unattainable.

In that case, only one price could prevail in the market, and two types of equilibria are possible: either

1. the price per unit of coverage equals the loss probability of high risks, π_H, high risks buy full insurance and low risks do not insure at all; or
2. the price lies between π_L and π_H, so that high risks are subsidized and buy more than 100% coverage, whereas low risks buy less than full insurance. Of course, regulation could enforce a limit of 100 percent on total coverage, which would prevent greater than full coverage among high risks.

In practice, health insurance companies try to determine total coverage by asking insureds about other coverage—but they cannot do so with perfect accuracy. Electronic methods for payments of claims to medical providers prevent duplicate benefits. However, there are insurers in the US that offer insurance that pays cash, based on the fact of a hospitalization or the onset of a disease like cancer, regardless of what other insurers pay for care. In practice, the ability to observe total coverage may be less relevant for preventing adverse selection than it is for controlling moral hazard for which incomplete coverage (i.e. co-payments) is crucial.

2.2.5. Equilibrium when Other Insurance Demand Determinants or Competition Varies

Insurance demand depends on more than risk; it also depends on risk aversion, price (premium or loading), and extent of moral hazard. Does the market equilibrium and attendant possibility of adverse selection vary in theory with any of these factors?

The general message is that as long as the market is competitive and any such other factors are distributed independently of risk, their variation does not matter. Chiappori et al. (2006) show rigorously that the positive correlation property (risk correlated with amount of coverage) holds under a wide variety of assumptions as long as insurers are free to offer any coverage they like and competition constrains economic profits. Even if insurers have knowledge of these other demand determinants (e.g. know who is more risk-averse and therefore more willing to pay for insurance), competition prevents them from taking advantage of any such information. The only examples of adverse selection in competitive health insurance markets are in markets where coverage is relatively unimportant, such as private insurance in the UK (Vera-Hernández, 2003).

Things potentially change if insurers have market power. Then they will be expected to discriminate in price if they have information on demand determinants. Even if they do not, they will know that people with stronger risk aversion will demand more insurance at a given price and so will mark up premiums more for more generous coverage.

If the zero profit constraint continues to hold but not all levels of coverage are available (either because of marketing costs or regulation) that could limit the scope of adverse selection but is unlikely to matter significantly as long as offered options roughly span the set of choices.

In contrast, if other demand determinants are correlated with risk—for example, if either risk aversion or cognitive ability is higher for lower risks—then competitive equilibrium might display either no relationship between risk and coverage or even favorable selection, as noted earlier.

2.3. Variation in Lifetime Risk versus Single Year Risk: Implications for Insurance Design and Market Equilibria

2.3.1. Reclassification Risk

Health insurance is typically obtained over multiple time periods, and for any individual the risk level may change over time. Arrow (1963) has advised us that in such settings the ideal is "insurance with a longer time perspective," but what form might such insurance take?

Consider a simple but realistic model where a population begins as a set of homogeneous low risks but expects the onset of chronic conditions to convert some proportion each time period to higher risk status. If insurance is priced for finite short

time periods (e.g. a year) and is priced based on the risk level at the beginning of each period, those people who become high risks will find their future premiums higher as they are reclassified into higher risk classes at the beginning of the next policy period. Other things being equal, risk-averse people would prefer to protect themselves against unpredictable fluctuations in future premiums. Are there ways markets can and do provide such protection?

2.3.2. Insurance Features to Deal with Reclassification Risk: Guaranteed Renewability

The simplest solution would be to sell insurance with a single lifetime premium, paid in an initial period when most people will still be of equal risk. These premiums would be high enough to cover the expected present discounted value of lifetime medical costs, given expectations about the onset and medical (and non-medical) costs of uncertain shocks to health status. In principle, a binding contract obligating buyers to pay the premium in installments would also be possible. Of course, uncertainty about future medical prices and technology would make offering such insurance a challenge, and the possibility of insuring future cost levels (even for constant medical risk) is limited.

Such a long-term insurance contract might be limited both by capital market imperfections (cost of raising the upfront lifetime premium) and the difficulty of enforcing future behavior (by buyers or sellers). So is there an alternative market arrangement in which premiums are paid at shorter time intervals, but still without the need for binding contracts for buyers?[1] The answer is affirmative. Consider a simple three-period model in which a given proportion of low risks (with constant low loss probability) convert to high risks in each period, and then remain high risks until the end. (We use a three-period model to illustrate the time–age path of premiums.) Suppose an insurer offers a premium schedule for the three periods in which the last period premium is the low-risk premium, the second last period premium is the low-risk premium plus the difference between the high- and low-risk premiums for the proportion of the initial population who convert to high risk in that period, and the initial period premium is the second period premium plus the difference between the low- and high-risk premium times the proportion of the population who convert to high risk in the first period. (Note that the proportion of the initial population

[1] This raises the more fundamental issue of the relevant time period for insurance demand and evaluation. With perfect capital markets and perfect foresight the right perspective would be the lifetime perspective, but it is clear that even well-off consumers do not want to finance a high medical cost in the current period by spreading it over time (even if they could do so). On the other hand, no one would want to plan insurance based on the risk of a high expense relative to income in a single day. Ehrlich and Becker (1972) distinguish self insurance (financing by borrowing or drawing down savings) and market insurance as two alternatives. How these fit together when risk varies over periods shorter than lifetime is unknown.

converting to high risk is lower in the second period than in the first because some people have already converted to high risk.)

It is easy to see that this schedule will be attractive to everyone, regardless of risk level, in all periods, since it guarantees premiums and since no subset of risk can do better by dropping out in any period. This "original guaranteed renewable" (OGR) premium concept was developed independently by Cochrane (1995) and Pauly et al. (1995). In effect, it is an arrangement in which the total premium in each period but the last has two parts, one part to cover the unexpected medical spending that may occur in that period, and the other part to cover the (present discounted value of) the increment in future premiums due to the risk of becoming a high risk in that period. It is incentive compatible and avoids both the risk of reclassification and (if it might have occurred) adverse selection. The premium still rises with overall health care costs.

Some features of this model are deserving of comment. If the low-risk premium is constant over time, the OGR premium schedule will be one in which premiums eventually decline over time as the extent of "frontloading" to protect those who become high risks diminishes. Empirically, because expected costs of a given benefit package even for low risks increase with age, the actual time path of GR insurance will be one in which premiums rise with age—but less steeply than if there were single period risk rating.

While low risks should rationally remain with the initial seller, it does not matter if they switch sellers, since the prepayments needed to assure low and stable premiums for those who have become high risks have already been collected. In this sense, this arrangement is not vulnerable to turnover of low risks. But if high risks expect to leave this firm for something more attractive (a public program or labor-related benefits), this diminishes the appeal of the GR arrangement.

One potential problem with this arrangement concerns the possibility that the insurer may not discharge the full contract for the high risks; it might compromise in some way on quality or service (however that might be defined) because the high risks have no attractive outside alternative (van de Ven and van Vliet, 1992), or it might try, despite the contract provision, to charge a selectively higher premium to the high risks, or just charge higher premiums to everyone. Usually the insurance contract is not explicit about the method of determining future premiums or the guarantees of all dimensions of future quality or service. Or the insurer might raise premiums for the class while at the same time offering a low premium to those who remain low risks if they leave the class. Reputation effects should inhibit such behavior (if I am still a low risk, but worried about the future, why would I sign up for coverage in a new class from a firm that has just increased its premiums to high risks?), but may not prevent it. It is in principle possible to design a more complex contract that gives those who have become high risks and who wish to change insurers the right to claim

the frontloading to which they would be entitled if they remained with the original seller. Cochrane (1995) imagines that high risks could demand a "dividend" which would be enough to cover high-risk premiums at other sellers, thus diminishing the original insurer's incentive to skimp. (Note that if this provision is in place then there should be little reason for insureds to actually leave; with proper incentives for quality, they should be willing (if not happy) to stay "married to their insurer.")

The other potential problem concerns the time path of premiums. If capital markets were perfect, consumers could pay a lifetime premium and there would be no need for guaranteed renewability. But while OGR reduces the capital burden for insurance early in life, the high frontloading in principle might still be a problem. If it were, a potential solution would be to deviate from the OGR model in which all risks pay the same premium in any time period, and instead allow some modest risk rating with attendant modest reclassification risk as a way of reducing the amount of premium that is frontloaded. However, as we shall see, for health insurance the frontloading burden may not be so large because the low-risk premium will increase with age and with medical progress.

2.3.3. Empirical Evidence on Guaranteed Renewability in Individual Insurance Markets and in Mixed Markets

2.3.3.1. United States

The most surprising empirical fact here is historical. Individual health insurance in the United States typically included a GR provision even before it was required by law in 1998. That is, in response to market demand, individual insurers usually promised not to re-underwrite (except for specifically designated temporary insurance), so that a person would not be singled out for a premium increase because of claims experience. In this way it was following on similar provisions in disability insurance and term life insurance. While information is imperfect, it appears that, although this provision was not universal and was sometimes violated in practice (often by reviewing the person's initial application and "rescinding" a low-risk classification if there was error or a possibility of error), individual insurance did display two characteristics consistent with GR frontloading, premium to claims ratio higher in the early years of coverage and relative insensitivity of premiums to risk (Pauly and Herring, 1999).

More recent work (Herring and Pauly, 2006) showed that the actual time path for OGR coverage, given the relationship between age and risk in the United States, is one in which the premium is actually lower at younger ages than at older ones, though not as low as under single period risk rating. "Risk" is defined as expected expenses (and so is correlated with but is not identical to measures of health or health status, whether physiological or subjective). Compared to the age profile based on average risk at each age, OGR is somewhat higher at younger ages but lower at older ages. The amount of "excess" frontloading is 34 percent of the (modest) healthy

young person premium, high enough to matter but not so high as to present a high barrier to financing. Moreover, they found that the actual age profile of premiums for individual health insurance in the US seems close to the OGR profile, especially after adjusting for (involuntary) turnover, further evidence that the market actually does what it promises.

Despite some cases in which insurers avoided or shed high risks under GR, the protection provided by GR is substantial. Compared to small group health insurance (where there is no protection against re-underwriting at the individual level), high risks were much less likely to lose coverage, other things equal. Thus it is far from obvious that group insurance for a person who might leave a job provides better protection against changes over time in health insurance risk than does individual insurance.

2.3.3.2. Germany

Germany is the OECD country outside the US with the largest private health insurance market. Private insurance in Germany primarily serves the self-employed, civil servants, and upper-income employees who choose private insurance over social insurance. By law, private insurance premiums must be calculated by the principle that they stay constant during the whole lifetime of the customer as long as the overall level of medical technology and prices does not change. This means that the customer is insured not only: (1) against health expenditures in the current period; but also (2) against changes of his individual risk assessment (reclassification risk); and (3) against rising health expenditures with age, whereas he is not insured against the (systemic) risks of medical progress and health care cost inflation. Premiums thus only depend on gender and age of entry into the contract (and, possibly, supplements for excess risk at entry). As a counterpart, the insurance company must display so-called "aging provisions" in its balance sheet, which account for the gap between expected future health care expenditures of and future premium revenues from its present set of customers. The whole arrangement is a combination of GR and a savings process which helps finance the predictable increase of health expenditures with age, in the absence of reclassification.

The main problem with this arrangement is that it may be possible to distinguish the premium which covers type (1) of risk from the accrual to the aging provision, but it is very difficult to disentangle the premium for the coverage of reclassification risk (2) from the "pure savings" part of the premium (3). Hence, when the insured wants to cancel his insurance contract and switch to another insurer, it is not clear which part of the aging provision should be transferable (Baumann et al., 2008). In fact, German law stipulated until 2008 that aging provision could not be transferred at all, and thus there was practically no competition between private health insurers except for first-time customers. The law was changed in 2009 so that now part of the

aging provisions has to be transferred to the new insurer in case of a move from one company to another, but the method of calculation is simply dividing the total amount of aging provisions in a demographic group (defined by age, sex, and age of entry) by the number of heads in that group. Instead, the correct procedure would be to calculate and transfer "individualized prospective aging provisions" by the method described above (expected future health care expenditures minus expected future premium revenue; see Cochrane, 1995).

3. GROUP INSURANCE

Employment-based group insurance is the predominant form of private health insurance in the United States, and it is also present in many other countries. In some countries, such as the US, employer-sponsored coverage is explicitly promoted through tax subsidies. While subsidization promotes employment-based coverage, it is "neither necessary nor sufficient" to explain the prominence and durability of voluntary group coverage more generally (Glied, 2005). In the US, employers often provided health insurance prior to the implementation of policies favoring employer-sponsored coverage over individual purchase. In addition, voluntary employment-based coverage exists in many other countries, even when tax subsidies are not available (i.e. UK supplemental market). Thus, employment-based group coverage may emerge even in the absence of policies promoting its existence, perhaps as a response to inefficiency in individual markets. Because policy often promotes employment-based group purchasing, however, it is difficult to disentangle the extent to which group purchasing represents the emergence of an alternative institutional arrangement for purchasing coverage in response to inefficiency in the individual market from an apparent normative preference of policy makers in promoting this institutional arrangement.

Despite the prominence of group purchasing, its economics, particularly with respect to the implications of group purchase for the performance of insurance markets in response to risk variation, has received relatively little formal attention among health economists.

3.1. Group Insurance Model and Theory

Several characteristics define group insurance:

1. Eligibility determined by worker status (usually as employee although there are some guilds of workers (often self-employed) who obtain insurance in a group, such as the Screen Actors Guild or the carpenters' or teamsters' unions);
2. Choice of number and types of insurance policies determined by the employer or group leadership;

3. Group membership and whether to enroll in the coverage offered by the group are both voluntary.

Group purchasing has two advantages over individual purchasing. First, groups experience less stringent underwriting than individuals. Usually a large share of the premium is covered as part of compensation, with only at most a minority of the premium paid by the insured. These provisions mean that the great bulk of workers take insurance through the job, and the premium is tied to the claims experience of the group; as a result, outside insurers usually do not find it worthwhile to risk underwrite or screen applicants to group insurance. Moreover, the set of workers can change over the period of coverage without affecting the premium charged. This means that, as long as the worker stays employed at the firm, changes in risk level associated with chronic conditions will not lead to changes in premiums the worker pays: there is protection against risk reclassification. (Of course, this protection is lost should the worker lose or change jobs.) Second, group coverage has lower administrative costs than individual insurance, largely because it is not necessary to persuade individual workers to take coverage; the effective subsidy to each individual's decision about coverage greatly reduces administrative costs in its own right.

In this type of group purchasing, risk variation may affect worker behavior in two ways. First, individual risk may influence the choice of employment and corresponding employer-provided benefit packages. Individual risk may also influence the choice of whether and in which plan to enroll within a firm. In other words, there can be risk selection both across and within firms.

3.2. Group Insurance and Current-period Risk

Two features of group insurance make it similar in theory to a local public good—the ability of workers to move across firms and the role of the employer (or group leader) in choosing a single or subset of potential health insurance plans of offer to workers. Analyzing group coverage from this perspective, Goldstein and Pauly (1976) demonstrate that, when employees vary in their preferences for health insurance, but are similar along other dimensions including their risk, labor market competition will lead to perfect sorting of workers across firms, with each firm offering a single plan which represents the most preferred plan for a particular subgroup of workers. This model highlights two important mechanisms which influence the functioning of group markets. First, cost-minimizing employers have strong incentives to offer workers efficient levels of coverage, where efficient is defined as the trade-off between health insurance and cash wages the worker would have chosen in the absence of group coverage. Second, labor market mobility tends to limit the degree to which outcomes in the employer-sponsored market in any one firm can deviate from those at other firms or

in individual markets. Using the Goldstein and Pauly framework as a starting point, we consider the impact of risk variation on worker sorting across and within firms in a single period model.

3.2.1. Worker Sorting and Wage Incidence

Generally the incidence of insurance costs nominally "contributed" by the employer will be on money wages. In the Goldstein and Pauly model above, workers were homogeneous with respect to risk, and wage offsets varied across but not within firms based on plan generosity. When the labor force consists of workers with heterogeneous risks, the extent to which wage offsets vary with individual risk will affect the stability of group markets. If we consider the simple case in which the entire premium is taken from compensation, and if we assume that money wages are reduced uniformly for all workers in a firm regardless of risk level, there could be substantial adverse selection as high-risk workers avoid firms not offering benefits and select employment at firms offering more generous benefits. While low-risk workers would have incentives to separate into homogeneous firms with lower than average premiums and correspondingly higher wages, high-risk workers would seek to join low-risk firms for the same reasons, creating dynamics similar to those in the standard Rothschild and Stiglitz model. If we instead assume that wage offsets vary with individual current-period risk, then, maintaining the assumption of perfect labor mobility, we return to the local public good equilibrium: workers sort among firms based on their preferences for health insurance. In this situation, in contrast, worker wages reflect individual risk due to differential wage offsets for health insurance. In short, movements away from risk-based individual incidence within employment-based groups create greater pressure for risk selection at the group level.

There is evidence that some dimensions of risk do affect the incidence of premiums on wages, so that higher-risk workers' wages are reduced more than lower-risk workers' wages when both are in the same insurance plan. For this reason, the apparently uniform explicit premiums in group insurance do not necessarily mean there is no risk of discrimination. Obesity, older age, and child-bearing status (which increase risk, other things being equal) have been shown to affect wages with group insurance (and to a greater extent than any effect on wages in a firm not offering insurance coverage) (Bhattacharya and Bundorf, 2009; Pauly and Herring, 1999; Gruber, 1994). But there are many other characteristics which are likely not reflected in wage offsets. As in standard models of insurance equilibrium in the presence of asymmetric information, this unpriced risk variation may cause adverse selection. In employment-based group markets, however, the adverse selection problem from the health insurance market "spills over" into the labor market (Bhattacharya and Vogt, 2006).

Some qualifications to these simple observations are as follows: given the small fraction of the premium explicitly charged for the lowest premium option, there will ordinarily not be adverse selection between individual and group insurance, because even a risk-rated, low-risk premium in individual insurance will usually be *greater* than the explicit worker premium in group insurance. So even low-risk workers with group insurance will not move to individual insurance. Workers may sort among firms based on risk, though we have little evidence or theory for this case. For there to be adverse selection between individual and group insurance, low-risk workers would need to be able to recover both their contribution and the employer contribution if they opted for risk-rated individual coverage, and usually there is no mechanism by which the employer share can be captured by individual workers who decide not to take group coverage.

Despite the likely absence of perfect individual risk rating through wage offsets, employment-based group purchasing appears to provide a relatively stable solution in many settings. What factors account for this apparent stability? Labor market friction likely plays a role. Workers come together in firms for purposes other than obtaining insurance, and this alternative basis for grouping (and the search and switching costs associated with different jobs) is a barrier against frictionless adverse selection. Consistent with this explanation, Bhattacharya and Vogt (2006) find that industries in which jobs typically require greater job-specific human capital have higher rates of insurance coverage. In addition, selection effects may be limited by the need for employers to design compensation packages which attract adequate numbers of average-risk workers. It is hard to think of a model in which benefits attract the average-risk worker enough to matter (in lowering employer total compensation cost) but do not differentially appeal to higher-risk workers. In the US, the extent to which risk-based selection is problematic in group markets is also likely tempered by the existence of public programs, including Medicaid and Medicare, which serve disproportionately high-risk populations (Cogan et al., 2010), and the tax subsidy to employment-based insurance which encourages low-risk workers to remain within the group market by restricting subsidies to this setting (Selden, 1999). Finally, some adverse selection may exist in the employer-sponsored market due to the absence of individual risk rating of premiums. Bundorf et al. (2010) document that, among people potentially eligible for employer-sponsored coverage based on family employment status, rates of insurance coverage increase with risk, particularly for people with low to medium incomes.

3.2.2. Premium Differentials, Efficiency, and Firm Objectives

Risk variation may also affect employee behavior within firms. Many employers offer multiple plans potentially leading to risk selection across plans if higher-risk workers choose more generous coverage. Risk selection may even occur in firms offering a

single plan if the plan requires an explicit premium contribution and low-risk employees are more likely to opt out of coverage in response to that contribution.

The potential for risk selection within the firm raises questions of why firms offer multiple plans and require employee premium contributions. Theoretical research has focused primarily on heterogeneity in employee preferences for coverage as an explanation for choice and less on the implications of heterogeneity in risk. Goldstein and Pauly introduce the potential for firms composed of employees with heterogeneous preferences for coverage into the local public good model by relaxing the assumption that workers are perfectly substitutable and instead assume that firms require workers of differing types for production purposes and that preferences for health insurance vary across worker types. They show that, for an employer offering a single plan, the cost-minimizing plan reflects a weighted average of the preferences of different types of workers. Firms may instead offer multiple plans in response to heterogeneous preferences of employees, balancing the trade-off between the greater administrative costs associated with managing multiple plans and the savings associated with greater customization with respect to employee preferences (Bundorf, 2002). While offering multiple plans may alternatively be a strategy by employers to promote greater within-firm competition among plans for enrollees (Enthoven and Kronick, 1989), Bundorf (2010) demonstrates that patterns of premiums and coverage are more consistent with choice as a response to diverse employee preferences than an explicit strategy to promote greater within-firm competition.

Offering more than one plan, however, creates the possibility for risk-based selection across plans. While employers often require different contributions for different plans, they rarely vary the contribution for a particular plan by individual characteristics related to risk. The widespread exception is family size—contributions often vary by the number of people covered under a policy—and sometimes contributions vary by employee wage. Although federal regulation currently prohibits varying contributions by age or health status, even when it was not forbidden by law, employee premiums almost never varied with employee risk. Thus, the potential for adverse selection is significant given that the prices faced by enrollees do not vary by easily observable characteristics related to risk. Cutler and Reber (1998) document a situation in which offering multiple plans and requiring employees to pay the full marginal premium for more expensive coverage led to significant adverse selection into more generous plans. If fostering/permitting adverse selection is viewed as (or is) a cost to employer profits, then there is a trade-off, as usual, between offering policies to fit differences in individual preferences that are unrelated to risk levels and causing adverse selection. The employer, however, can determine the size of the differential premium for one plan over another, and even when both plans are bought from outside insurers (rather than being self-insured), there is no necessity that the differential reflect the actual difference in average premiums.

Given the constraint on explicit discrimination across workers, there has been a large amount of work specifying what premium differentials should be (for second-best efficiency) and some work examining what those differentials are in reality. The model developed by Cutler and Reber (1998), Pauly and Herring (1999), and Cutler and Zeckhauser (2000) imagines that the goal of the group insurance plan is second-best optimality in employee choice of plan. Assuming that different plan designs yield higher benefits to workers at different risk levels, and that the high-risk plan is more costly than the low-risk plan, the problem is to incentivize workers to choose the appropriate matching plan. A strategy which sets a fixed dollar employer contribution and then sets the premium differential equal to the difference in average expenses of workers who choose each plan type will not be efficient, because the premium differential will reflect both the "true" incremental cost of the more generous plan for the high risks and the difference in risk levels. This will cause a death spiral for the more generous plan. If there are only two risk levels, setting the premium differential at the incremental cost for high risks will cause an efficient allocation of risks to plans, but if the distribution of risks is continuous the actual difference will reflect the average incremental cost, not the optimal incremental cost for the risk level at the margin of indifference between the two plans. Rather than asking what is optimal for the set of workers, Miller (2005) asks what contribution policy an employer minimizing total compensation costs would choose in a similar setting. He finds, not surprisingly, that such an employer will choose to set the monopoly price for some of the plans. He does not consider possible effects of employer premium-pricing policy on the distribution of workers across employers.

These models rely on two important assumptions: (1) that more generous coverage is efficient for high risks but not for low risks because they have stronger preferences for more generous plans; and (2) that preferences for coverage are perfectly correlated with risk type (Bundorf et al., 2008). More generally, a single uniform contribution in the presence of heterogeneous risks will achieve a first-best allocation if and only if the single crossing property holds—the value that consumers place on the more generous plan increases more quickly with risk than the incremental cost. In this situation, the key challenge for setting the socially efficient contribution is identifying which risk type is marginal. But it is entirely possible that the single crossing property does not hold. Perhaps less generous plans generate greater cost savings for high than low risks. In this case, preferences for more generous coverage may increase less quickly with risk than incremental costs and no single uniform contribution can generate an efficient allocation. More generally, preferences for different types of plans may be either uncorrelated or imperfectly correlated with risk. In this case, a uniform contribution will represent a second-best solution and the second-best contribution will depend upon the distribution of risk types and preferences for

coverage in the population. A first-best allocation would require risk-rated contributions.

Important gaps remain in research in this area. One is that each of these studies treats the composition of the group as fixed. A more realistic treatment would allow the mix of workers to depend on the plans offered and their premium differentials. Moreover, in reality employers usually do not charge as explicit premiums either the welfare maximizing configuration or the labor cost minimizing configuration for a given set of workers. Surprisingly little is known about what employers actually do. Either their objectives differ from these goals, or they have imperfect information on how to do so. Nevertheless, the conclusion is that it is hard to describe the positive empirical benchmark for what happens (in the absence of regulation) and therefore even harder to describe how things would change if external circumstances or incentives changed.

3.3. Group Insurance and Risk Variation Over Time

While things therefore are far from perfect, are there nevertheless ways in which group insurance improves how consumers are treated when it comes to risk? The most obvious reason for a positive answer is that the insensitivity of explicit premiums and money wages to risk constitutes a kind of informal mechanism for protecting people against the reclassification risk they might find in simple (if unrealistic) models of individual health insurance markets. In some settings, this may be a reason for the existence of group insurance, as an alternative and perhaps less costly arrangement than individual insurance with guaranteed renewability. That is, as long as the worker continues group coverage, the explicit premium he pays does not change even if risk changes for the worker or dependents. Bhattacharya and Vogt (2006) demonstrate that this type of risk pooling over time is dependent upon the degree to which employment transitions are costly and the extent to which health transitions are persistent. In contrast to the GR approach, there is no need to frontload the premiums to offset growth in risk should a firm's workforce age, because the explicit employee premium will almost always be less than the premium charged by any individual insurer. (There is a possibility of adverse selection from another employer who offered less generous coverage and higher money wages in order to skim off the healthy workers.)

But the protection provided by group insurance extends only as far as the individual firm employment contract. Let the high-risk worker change or lose her job, and she may be confronted with substantial increases in premiums or reductions in wages. This type of concern is consistent with the job lock literature which finds that people with employment-based health insurance, particularly those who are high risk, are less likely to transition jobs than those in jobs without such coverage (see Gruber, 2000,

for a review). Indeed, Pauly and Lieberthal (2008) find that, among those in poor health, the likelihood of losing health insurance is greater for those with coverage from a small firm than for those with individual coverage.

4. PUBLIC POLICY TOWARD RISK VARIATION

A striking feature of the health care systems of all developed countries is the prominence of policies governing the degree of variation in premiums in health insurance markets. In countries with single payer systems, risk variation in premiums is essentially eliminated by the absence of premium payments. In systems in which people make payments for health insurance, the degree to which they are related to individual risk is nearly always influenced by regulation.

Both efficiency and equity rationales for the pervasiveness of government intervention along this dimension exist. As discussed earlier, in theory health insurance markets may operate inefficiently due to information asymmetries between consumers and insurers, and government intervention is a potential response to this inefficiency. Most of the literature on the rationale for government intervention in insurance markets due to adverse selection, however, takes as its starting point a single-period model of individual purchase of coverage. Yet we have shown that the way in which unregulated private markets deal with risk is, because of guaranteed renewability and group insurance, a far cry from simple single-period risk rating. That is, the alternative to insurance markets with public intervention is not necessarily or even usually individual insurance markets of the Rothschild–Stiglitz type subject to adverse selection. Thus we face a serious dilemma at the outset. If regulation is already present and we are trying to see if it is justified, we really do not know what would happen if regulations and tax subsidies were removed, but we do know that it is unlikely to fit any of the theoretical models well. To evaluate the welfare effects of regulation, we need more complete knowledge of the nature of the market or markets that do exist in the absence of regulation.

There are some powerful reasons why we would not expect unregulated markets to fit the Rothschild–Stiglitz model. Arrow (1963) noted that the medical care sector as a whole as well as the health insurance sector display properties inconsistent with the description implied by simple models of competitive equilibrium because of a variety of differences between risky medical care and standard commodities. Some of those differences surely come from regulation that is universal (like physician licensure), but some also come from differences from the simple model apparently generated by the market itself to deal with what would otherwise be uninsured risk: the

most prominent of these are guaranteed renewability and group insurance. That is, when we take the fetters off health insurance to generate an unfettered free market, we do not get the simple Rothschild–Stiglitz model. While both guaranteed renewability and group insurance historically were in the US accompanied by and doubtless fostered by favorable tax and regulatory treatment, it remains true that it is difficult if not impossible to find an empirical counterpart to the simple single-period risk rating insurer market equilibrium even in a setting where regulation of individual coverage is minimal and tax subsidies small. Both public and private sectors, probably to avoid market failure, have set up markets different from the simple model. It seems plausible, as Arrow argued, that citizens will develop multiple ways, some public and some private, to deal with the peculiarities and inefficiency of the health insurance system. But then one will logically be unable to find the simple but inefficient model in practice. There still may be some variation (driven by heterogeneity in individual behavior in response to insurance market inefficiencies) and some inefficiency in the real world since individual responses will represent a second-best solution.

While there is a possible if unproven efficiency argument based on reclassification risk against a market with single-period risk-based premiums, the strongest arguments among policy makers (in contrast to economists) for regulation to inhibit risk-based pricing are based either on equity or some other postulated social norms. Presumably the reason for policies which limit the degree of risk-based premium variation is the expectation that, in their absence, premiums would vary with risk in undesired ways. On equity grounds, charging high risks more may be said to be unjust, although a full treatment of equity should probably account for income as well as risk variation. Or sometimes it is simply asserted that "society" has decided to "pool risks." Insurance with risk rating in a single-period setting attains the optimal amount of risk pooling; presumably this is a statement calling for risk averaging rather than the pooling of risk of any event that is yet to happen within the term of insurance coverage. These arguments can be addressed against risk-based pricing of a high-risk insured new to a given insurer even when that person had or could have had premium protection through guaranteed renewability in an earlier time period before becoming a high risk.

Policies which limit premium variation to achieve equity objectives, however, generate economic inefficiency. From this perspective, the key economic question is how to achieve the normatively optimal distribution as potential policy interventions generate different trade-offs between efficiency and equity. Once again, however, the lack of a positive, second-best benchmark restricts our ability to evaluate these trade-offs.

Despite reasons to be less concerned about the way unregulated markets handle risk, the possibilities of inequity and inefficiency have motivated regulation. We therefore explore what impact such regulation might have. In the following sections, we

discuss the types of policy instruments that have been analyzed in theory to address each rationale. We think it is worth analyzing these policies even while retaining substantial skepticism about the relevance of the rationale.

4.1. Regulation for Efficiency in Markets with Risk Variation

4.1.1. Policies to Address Inefficiency due to Anticipated Asymmetric Information and Associated Adverse Selection

One of the efficiency justifications of public intervention in health insurance markets is based on the theoretical result that a Rothschild–Stiglitz equilibrium is not Pareto efficient if the share of high risks is moderately large. (The welfare evaluation of cases in which no equilibrium exists is less clear, though one can certainly assert in such cases that no efficient equilibrium exists.) In this case a combination of mandatory partial insurance with community rating and voluntary supplementary insurance can make both risk types better off (Eckstein et al., 1985). Even this is only a second-best optimum because in a first-best world—and in the absence of moral hazard—every risk-averse individual would want to buy full coverage at a fair premium. In contrast, in the equilibrium of the mixed public–private insurance market, low risks are still underinsured and pay a total price that exceeds the fair premium.

This simple efficiency justification of mandatory public insurance breaks down because unregulated markets for health insurance are better described by alternative models. For example, if the assumptions of the Wilson–Spence–Miyazaki model hold, mandatory public insurance no longer increases efficiency relative to the market equilibrium because the corresponding equilibrium in that model is always Pareto optimal given the informational constraints (Crocker and Snow, 1985). Perhaps more realistically, as discussed earlier, it breaks down even further if insurance is offered in a group setting or is offered as individual coverage with guaranteed renewability to a population with homogeneous risks (that change differentially over time).

4.1.2. Administrative Costs Associated with Risk Variation and Methods of Reducing Them

Underwriting costs must be incurred if risks vary and premiums depend on risk. Explicit underwriting costs do not seem to be large, perhaps at most 1 percent of premiums, but they are not negligible. And some of the selling expense may be attributable to underwriting if insurers pay brokers to avoid or at least identify higher risks. How much underwriting expense an insurer will choose to incur depends on the extent to which risk actually varies in its population of potential insured. In a simple model where risks are either "high" or "low," the expense of distinguishing between these two risks will depend on their proportion in the population and the difference in expected damages between the two classes. The obvious point is that if risk does not vary much it may not pay an insurer to incur much expense to

discriminate among risks. It also will not pay to incur high costs of risk discrimination if either type of risk is a small fraction of applicants for the insurer. If there are few high risks it will not pay to find them, and if there are few low risks they will not pay to have this found out. The fact that underwriting does occur, however, suggests that the benefits to health insurers of discriminating among risk types, at least to some degree, exceed the costs.

Regulations requiring community rating will reduce explicit underwriting costs. Community rating, however, introduces the need for a third party to implement a system and transfers across insurers based on enrollee risk in order to prevent insurers from competing on risk selection. Thus, a reduction in the cost of underwriting on the part of insurers is offset by the introduction of resources devoted to risk adjustment. Community rating has additional efficiency implications which we consider in sections 4.3.4 and 4.3.5.

4.2. Regulation for Equity in Insurance Markets with Risk Variation

Private markets may function efficiently but the market outcome may deviate from society's normative preferences for distribution, based on either health status or income, and government intervention may be an attempt to achieve this desired redistribution. In a private market for health insurance, the buyer seeking coverage from a new insurer would pay a premium which—given the amount of coverage—reflects the expected value of future health care costs. Those with a higher risk of illness would thus have to pay higher premiums than those with a lower risk. Many consider this market solution unjust—that those unlucky enough to be in poor health must pay more for access to health care, particularly when poor health is due to inequality in health endowments (e.g. congenital diseases or handicaps) rather than differences in the person's health-related behavior (lung cancer caused by smoking). Risk rating of the former is often considered less acceptable from a normative perspective than risk rating of the latter.

It is also important conceptually to distinguish between the single period and lifetime models when considering the rationale for policies to reduce risk variation in premiums. In particular, the single-period model introduces the risk of risk reclassification, which may be viewed as undesirable for either equity or efficiency reasons. As discussed earlier, insurance policies with guaranteed renewability can be an appropriate remedy for reclassification risk. Even long-term insurance contracts with guaranteed renewability, however, do not eliminate risk-based premium variation if differences in risk are present at the initiation of the contract. In this case, transfers from low to high risks can be justified using the concept of the veil of ignorance introduced by Harsanyi (1955) and Rawls (1971). People behind the veil of ignorance do not know whether they will be born as high or low risks. On top of their health

risk as such, they face a lottery concerning their expected health care costs and thus their health insurance premiums. Given risk aversion behind the veil of ignorance, they would be willing to insure against this premium risk. Hence, if the hypothetical veil of ignorance is accepted as the basis of a fair decision, not guided by particular interests, equalization of health insurance premiums between high and low risks may be desirable.

Given the present state of scientific knowledge, the proportion of the population who are high risks at birth is small, perhaps 1 to 4 percent. Specific public programs for the disabled as part of social insurance are able to address much of this type of concern. Provisions in unregulated insurance markets in the United States also provide substantial protection. If prospective parents have family insurance coverage, premiums for that coverage are generally not altered when a baby is added and, even if premiums are changed, the adjustment only takes account of the number of children put on the policy, not their health state. Thus, at least up until the child goes off the parents' coverage (now age 26), the family can be protected against the costs associated with a high-risk child.

While it is possible that genetic testing will eventually provide information on differences in lifetime total medical spending, at present low-cost genetic tests are sufficiently accurate for only a few rare conditions such as cystic fibrosis and Huntington's disease. If genetic tests become cheaper and more accurate over time, individuals should be advised by public authorities to purchase their desired amount of (life and health) insurance coverage before they undergo such a test to preserve their ignorance at the point of insurance purchasing. This would be a much milder public intervention than the one proposed by Tabarrok (1994), who proposes that genetic tests may only be sold to persons who can prove that they have already purchased "adequate" insurance coverage. Alternatively, the use of genetic test results for underwriting can be prohibited by law, which is done in many countries, but this regulation is not welfare enhancing because it creates adverse selection (Strohmenger and Wambach, 2000) as buyers of insurance use their genetic test information to decide on insurance coverage.

4.3. Government Intervention in Practice

The set of policies in place which influence the degree of premium variation vary significantly both across and within country. In theory, policy makers could achieve a desired final distribution of insurance, medical care, and income through a variety of mechanisms. For example, ignoring administrative costs, a system of community rating of insurance premiums combined with an individual mandate and income-based subsidies could achieve a similar outcome as a system of single-period risk rating combined with risk and income-based transfers. Yet the implications for efficiency of the different choices may differ substantially depending on the degree of price responsiveness among consumers, the distribution of risks within the population, and the

efficiency of the tax system. Thus, differences in policies may be driven by either the administrative or the efficiency costs of different mechanisms or the desired final distribution across countries. In practice, however, it is difficult to reconcile observed policies with economic criteria of either efficiency or equity. As discussed earlier, in the case of efficiency, the lack of understanding of the competitive benchmark hinders our ability to determine whether regulation is welfare improving. And in the case of an equity rationale, the normative criterion which policy makers are trying to achieve is rarely clearly expressed, making it difficult to determine whether observed policies are meeting distributional objectives and to evaluate the efficiency trade-off necessary to achieve that standard. Perhaps the most than can be said is that policy makers seem to have a strong normative preference against single-period risk-based pricing which appears to motivate substantial government intervention in health insurance markets.

In the next sections, we identify the different types of actual and proposed policies which influence risk variation in premiums and discuss some of the equity and efficiency implications.

4.3.1. Single Payer System

This solution has been chosen by countries such as the United Kingdom and Italy. By eliminating competition among insurers, a single payer necessarily eliminates the possibility of inefficiency due to adverse selection, and tax-based financing eliminates the explicit link between a person's health status and his or her contribution for health insurance. The efficiency costs, however, may be substantial. The elimination of competitive insurance markets eliminates any benefits from insurer competition in the form of cost, quality, and customization of insured medical care. Tax-based financing, the magnitude of which is substantial in a single payer system, also creates economic distortions.

4.3.2. Person-specific Transfers to High Risks

An alternative is to allow insurers to charge risk-based premiums and to delegate the task of subsidizing high risks to the tax-transfer system. For example, Pauly et al. (1992) propose introducing refundable tax credits which reflect a household's risk category and are inversely related to household income. Citizens with little or no tax liability would receive a transfer payment.[2] Calculating risk-based subsidies, however, may represent a challenge to a regulator who does not directly observe the health status of any given individual. In practice, only a few diseases which are easily diagnosed would probably be used to determine the transfers.

One approach would be to use the premium charged by a risk-rating insurer as an indicator of risk. Zweifel and Breuer (2006) propose to subsidize risk-based premiums

[2] See also van de Ven and Ellis (2000) for a discussion of this approach.

based on the extent to which they exceed a certain percentage of household income. In this way, they aim to target subsidies to individuals who are both high risk and low income, arguing that high risks with high incomes need not be subsidized. Zweifel and Breuer claim that in the presence of moral hazard, efficiency requires tailor-made individual insurance contracts in which individual parameters such as price elasticities play a role and that risk-based premium reductions for cost sharing require risk-based premiums to begin with. In their view, risk adjustment, in contrast, must always be imperfect and can never completely remove incentives for risk selection.

However, premium subsidies as envisaged by Zweifel and Breuer create a number of problems. First, they create an incentive for both the insurer and the insured to include additional services in the contract in order to increase the transfer. This can only be avoided by defining a detailed (minimum) benefit package. Then, however, an important advantage of competition in the health insurance market—diversity of insurance contracts—would be limited. Alternatively, the subsidy could be scaled not to the individual premium but to an average defined over a broadly defined risk class.

The second problem is that recipients of the transfer lose all their incentives to shop around for an inexpensive insurer because effectively their contribution is limited by a certain percentage of their income (van de Ven, 2006).

Kifmann and Roeder (2010) have analyzed under what conditions risk-based premiums accompanied by premium subsidies for low-income households are more "equitable" than community rating, where equity is measured by a Rawlsian social welfare function. They show that this is more likely the case the more negative is the correlation between health and productivity (earnings potential). In light of the vast literature on the association between income and health (Wagstaff and van Doorslaer, 2000), a negative correlation must be considered an exception rather than a rule. But even in this unlikely case, a combination of community rating and premium subsidies for low-income households would dominate the Zweifel–Breuer solution.

4.3.3. Person-specific Insurance Programs for High Risks

Another way to limit the degree of risk-based variation in insurance markets is to develop public programs targeted to high risks. These programs can take the form of public insurance with eligibility restricted to high risks or private insurance offered through subsidized high-risk pools at a price higher than that available to medium or low risks but lower than that available to high risks in the unsubsidized market. These pools accept people who have been deemed high risks by insurers; they charge premiums somewhat higher than standard and offer rather limited coverage in order to discourage people from being eager to enter the high-risk pool to claim high subsidies. And they only limit the coverage sought by those who want to use the pools; people who buy their insurance in the risk-rated open market may obtain whatever coverage they wish from competitive insurers. Even within the pool insurers can

compete on the basis of premiums since the per-person subsidy is usually specified as a predetermined amount or subject to a predetermined cap. The main problem such pools have had arises from limitations on financing by the state governments who, until health reform, organized and financed them; there were often waiting lists for the pools. A second problem with the state-run pools is that they were largely financed with assessments on other insurance purchases, thus increasing incentives for adverse selection there. The recently passed legislation reforming the US health care system establishes high-risk pools as a temporary measure to provide coverage to high risks until more extensive reform is implemented. The version in health reform is, however, funded by general revenue taxation (but with a relatively modest total budget).

In addition to reducing premiums for high risks who are made eligible for the coverage, these types of programs may make private coverage in regular markets more accessible for those high risks who are not eligible for the programs. Newhouse (1996) demonstrates how costly contracting can increase the likelihood of a pooling equilibrium in which insurers offer the more limited contract preferred by low risks. An implication of the Newhouse model is that a decrease in the size of the high-risk population increases the likelihood of a pooling equilibrium and increases the generosity of the coverage offered. In other words, when competitive markets are characterized by the assumptions of the Newhouse model, programs which remove a portion of the high risks from the competitive market may correct some of the distortions generated by asymmetric information. Cogan et al. (2010) demonstrate that the extension of Medicare coverage to the disabled in the 1970s had these types of effects.

4.3.4. Insurance Market Regulation

Many countries rely on private insurance but regulate the degree to which premiums may vary with risk. The concern, of course, is that, taken in isolation, this type of regulation may actually cause adverse selection. A number of studies have examined the effect of rate regulation on adverse selection in individual markets for both primary and supplemental health insurance. These studies compare rates of insurance coverage in states with laws restricting the extent to which insurers are allowed to vary premiums based on individual characteristics to rates in states without these types of rating restrictions. Most studies find that, while these laws have a relatively small effect on overall rates of coverage, the lack of an overall effect represents the combination of an increase in coverage among high risks and a reduction in coverage among low risks. For example, Davidoff et al. (2005) find that reforms in the small group market which restricted rating variance increased rates of coverage among high risks by 4.5 percentage points and reduced rates of coverage by 1.7 percentage points among low risks. This finding is remarkably consistent across settings, including the market for individual primary coverage (Herring and Pauly, 2007; Lo Sasso and Lurie, 2009) and private supplemental coverage (Bundorf and Simon, 2006). Researchers

have also documented a similar effect in the US small group market (Buchmueller and DiNardo, 2002; Monheit and Steinberg Schone, 2004; Simon, 2005; Davidoff et al., 2005). Buchmueller and DiNardo (2002) suggest that the lack of an overall effect on rates of coverage in the small group market was generated by a shift from indemnity to HMO coverage among low risks, and Lo Sasso and Lurie (2009) find a similar shift in response to rate regulation in the individual market. Thus, at best, restrictions in insurer rating practices generate a transfer between low and high risks. And in settings in which consumers are price responsive, rating restrictions could generate significant welfare loss due to adverse selection.

There is an idealized version of regulation that would deal optimally (in both efficiency and equity terms) with risk variation, but its requirements are very demanding. We outline this model primarily to show what compromises more realistic regulatory structures would have to make to achieve a second-best optimum. In particular, if the optimal insurance coverage for every person (taking into account variation in risk aversion, administrative loading, and income effects on the demand for medical care) has been determined, if insurers can be assumed to operate with maximum efficiency, if it is possible to adjust premiums for the ideal insurance for each person based on income and risk, and if purchase of the optimal insurance can be made mandatory with an effective mandate, then the outcome can be called ideal. Thus, in practice, using community rating to achieve a more equitable distribution of insurance requires mandatory purchase of coverage to keep low risks in the market and a system of risk adjustment combined with regulation of the benefit package to prevent insurers from distorting coverage for the purpose of risk selection. We discuss the challenges of risk adjustment in greater detail in section 4.3.5. This seems to be the conceptualization of regulation underlying the Dutch and to some extent the Swiss system, but it is hard to achieve in practice even there. It will be doubly hard to achieve in a more heterogeneous setting such as the United States.

4.3.5. Risk Adjustment

Risk adjustment, the process of administratively adjusting the premiums paid to health plans based on enrollee risk, is one piece of a regulatory response to the potential inefficiency generated by community rating requirements. Requiring insurers to accept any individual at a uniform premium leads to expected losses for high risks and expected profits for low risks, creating incentives for insurers to enroll low risks (*cream-skimming*) while avoiding high risks (*dumping*). Under competitive pressure, even not-for-profit insurers will have incentives to risk select since they will need a sufficient number of low risks to break even.

Risk selection can take two forms. On the one hand, health insurers can perform *direct risk selection* by influencing who signs a contract. Regulations requiring insurers to accept all applicants, known as guaranteed issue or open enrollment requirements,

are intended to address this type of behavior. Health plans, however, may use more subtle mechanisms than outright rejection to achieve favorable selection. For example, insurers may "lose" the application submitted by a person who is considered expensive and may encourage people who are expected to use little health care to enroll by offering them supplementary services at a discount or, in the extreme case, outright payments. Although there is little evidence of this type of direct risk selection in regulated markets in the US, an audit study found that German insurers who were unable to vary premiums based on geographical differences in expected expense were more likely to respond to applicants from low-cost than to those from high-cost areas (Bauhoff, 2010).

Indirect risk selection, on the other hand, refers to designing insurance products in order to induce consumers to self-select into different plans based on their risk. For example, plans with high cost sharing are more attractive to low risks who expect to use less care. Community rating is also generally accompanied by regulations specifying a minimum benefit package to limit the ability of insurers to risk select through benefit design. But even if the basic benefit package is fixed by the regulator, insurers can still engage in indirect risk selection. In markets in which private insurance supplements the public system, insurers may offer supplementary benefits to low risks at a discount, while charging high risks a mark-up if payments for the basic package are not properly risk adjusted (Kifmann, 2006). Competition among insurers translates this mark-up into a lower price for the basic benefit. As a consequence, cross-subsidies from low-risk to high-risk types are reduced, effectively undoing the cross-subsidization of community rating.

Managed care plans, however, may represent a greater challenge to a regulator concerned about indirect risk selection. Relative to traditional insurance coverage, managed care plans make greater use of supply-side cost sharing in order to create incentives for providers to deliver care more efficiently. Supply-side cost sharing, however, also creates incentives for risk selection on the part of providers, and thus generates a trade-off between efficiency in production and risk selection at the point of service (Newhouse, 1996). When health plans face strong incentives for risk selection, they may choose to design contracts with providers in ways which facilitate favorable selection at the expense of efficiency in care delivery. Because the techniques of managed care are not easily observable to a regulator, limiting inefficient behavior on the part of plans through regulation is difficult (Newhouse, 1996).

Theoretical studies have examined different ways in which managed care plans may distort coverage to achieve favorable selection including their choice of supply and demand size utilization controls (Eggleston, 2000), differential rationing across different types of services (Frank et al., 2000) and service-level carve-outs (Biglaiser and Ma, 2003). Testing the empirical predictions of the Frank et al. (2000) model using data from the US Medicare program, Cao and McGuire (2003) find evidence

consistent with HMOs choosing service levels to achieve favorable selection. Based on the same model, Ellis and McGuire (2007) develop an empirical index to identify which services are likely to be distorted by HMOs in order to achieve favorable selection. In their analysis, they introduce the distinction between predictability and predictiveness in health care spending. The utilization of a certain health care service is called *predictable* if a large part of the interpersonal variation in spending for this service can be predicted by using observable characteristics of the insured. The service is *predictive* if its utilization is highly correlated (positively or negatively) with the overall use of health care services by a particular patient. The authors then argue that services which are both predictable and predictive are candidates for rationing by managed care plans for the purpose of risk selection and construct an index which measures the potential for under- or overprovision. They propose that regulators could specify minimum levels of intensity for the services with particularly high values of the respective index in order to reduce the degree to which managed care plans distort services in response to incentives to risk select.

4.3.5.1. Defining Risk Adjustment

Risk adjustment is a regulatory response to the inefficiency created by restrictions in insurer rating practices which create incentives on the part of plans for risk selection. In this section, we define risk adjustment as the process of administratively setting premium payments to health plans based on the risk of enrollees. When regulation prevents health plans from charging consumers risk-rated premiums directly, risk adjustment essentially reintroduces risk-based variation in the payment a plan receives for enrolling a particular consumer.

Implementing risk adjustment requires a sponsor or an intermediary to collect revenues, either in the form of taxes or enrollee premium contributions, and to redistribute these revenues in the form of risk-based premium payments to health plans. To highlight the role of the intermediary, Figure 11.1 contrasts a competitive market with a regulated system relying on risk adjustment. In panel A, consumers purchase insurance directly from health plans and the premium the consumer pays is the premium which the plan receives for covering the consumer's medical expenditures. In panel B, in contrast, the payment made by a consumer differs from the premium received by the plan. In this section, we refer to the payment made by the consumer as the "enrollee contribution." As shown in panel B, enrollee contributions are often supplemented by tax-based revenue. We refer to the payment made by the intermediary to the health plan for providing covered services as the "premium payment." The key distinction between panels A and B for the purpose of risk adjustment is that the payment received by the plan is determined in the market in panel A and is determined administratively by the sponsor in panel B.

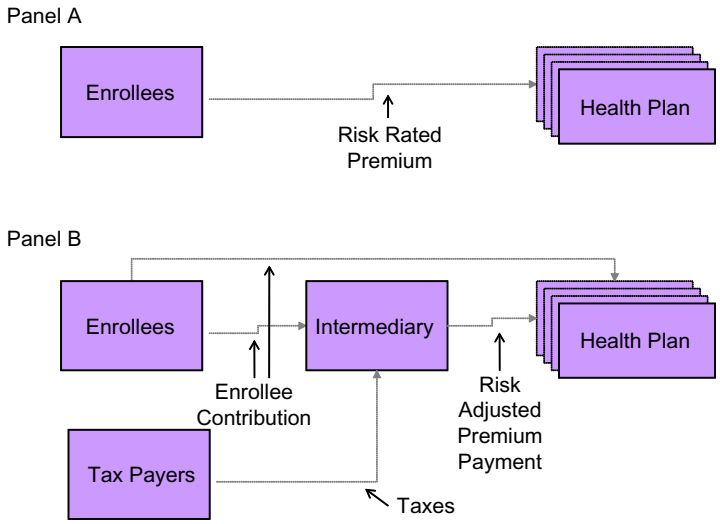

Figure 11.1 Competitive Market versus Regulated Market with Risk Adjustment.

We note that our definition of risk adjustment does not encompass administratively adjusting any component of the enrollee contribution. Keenan et al. (2001), for example, refer to the process of adjusting employee premium contributions to reflect differences in plan cost but not differences in enrollee risk as "risk adjustment of employee premium contributions." In other words, we define risk adjustment as a supply-side regulatory strategy. Other studies have referred to risk adjustment, as we have defined it here, as "risk equalization" (e.g. Armstrong et al., 2010).

Risk adjustment is used in a variety of settings, and, in practice, the simple model outlined above can be implemented in different ways. Van de Ven and Ellis (2000) differentiate between mandatory enrollee contributions which are not linked to plan choice and those which are. They refer to the former, which are often linked to income, as "solidarity contributions" and the latter as "premium contributions." Solidarity contributions are intended to reflect the redistributive objectives of the financing system, and the payment linked to plan choice is intended to create incentives for consumers to make efficient coverage choices.

Van de Ven and Ellis (2000) also describe two different modalities for the flow of funds for risk adjustment. In one, consumers make both a solidarity contribution to a sponsor and a premium contribution directly to a plan. In the other, consumers make a single contribution, including both solidarity and direct premium components, to health plans and then the sponsor administers a system of transfers across plans. The advantage of the latter model is that it reduces the extent of funds transfer across organizations and the associated administrative cost. The key feature which is common

across settings, however, is that an intermediary ultimately determines the degree to which the payment a plan receives for a particular enrollee varies with risk.

4.3.5.2. Risk Adjustment Methods

In practice, risk adjustment involves setting the premium payment to a plan proportional to the expected cost of a particular enrollee by weighting a base payment for different observable characteristics. Usually, the weights on the individual characteristics are developed from a regression estimating health care costs as a function of those characteristics. Research on the development of risk adjustment methods has focused primarily on how to maximize the explanatory power of the model used to derive these weights and which types of individual characteristics one would include in the model. Van de Ven and Ellis (2000) provide an excellent overview of the technical issues in estimating these models. Here we identify some of the issues in the debate over which characteristics one would include in a risk adjustment model. The simple risk adjustment model presented by Newhouse (1996) provides a framework for discussion:

$$Y_{i,t} = \alpha + X_{i,t}\beta + \mu_i + \varepsilon_{i,t}$$

where $Y_{i,t}$ is person i's medical care consumption in time period t, α is a constant, β is a coefficient, $X_{i,t}$ is a vector of risk adjusters, μ_i is a time-invariant, person-specific effect with a mean of zero, and $\varepsilon_{i,t}$ is a random error term with an expected value of zero.

Newhouse et al. (1989) raise the concern that a predictable component of person-specific variance, μ_i, exists even after controlling for observable characteristics in X such as age, health status and indicators of chronic conditions, suggesting that even detailed systems of risk adjustment may leave incentives for plans to risk select. They propose that adding a measure of prior health utilization can address this by increasing the predictive power of the model. Consistent with this concern, Hsu et al. (2009) demonstrate that measures of prior use significantly increase the predictive power of the relatively detailed risk adjustment formula used for the US Medicare Part D Prescription Drug Benefit and that premium payments based on this richer model reduce plan incentives for risk selection. Prior use, however, is also a measure of plan efficiency and conditioning payment on prior use dilutes plan incentives to produce efficiently. Thus, risk adjustment which conditions on prior use creates a trade-off between risk selection and efficiency in production of medical care (Newhouse, 1996). Alternative "risk-sharing" schemes exist and many are discussed by van de Ven and Ellis (2000). What they have in common is that the plan is reimbursed for some component of actual rather than predicted utilization, as a way of dealing with the likelihood that things may not always turn out as predicted.

Lamers et al. (2003) argue that a regulator would want to exclude some of the exogenous determinants of health care costs as risk adjusters based on equity criteria.[3] They view the risk-adjusted premium payment as a subsidy and the choice of which factors to include in X as a choice over which factors the regulator views as important to solidarity. For example, most systems of risk adjustment would include characteristics such as age, sex, and the presence of chronic conditions. But if income were strongly correlated with health care costs, would a regulator include income as a risk adjuster? Van de Ven and Ellis (2000) frame this largely as an equity issue and propose that a regulator would have to differentiate between the types of characteristics society would and would not like to subsidize. Yet if people with low incomes use more services, holding other measured factors such as health status constant, then risk adjustment which did not condition on income would create incentives for insurers to avoid people with low income. Hsu et al. (2010) provide an example of income-based selection in the form of competing plans avoiding high-cost, low-income beneficiaries due to the inability of the risk adjustment system used in the context of US Medicare Part D prescription drug coverage to account for differences in expected utilization by income.

Breyer et al. (2003) discovered three variables that added to measured R-squared in a regression explaining expenditures of a German sickness fund in 1993: income, being single, and being in the last year of life. Due to the income-related contributions to the German social health insurance system, sickness funds can observe their members' incomes. First, the authors found an income elasticity of health care expenditures of roughly -0.5. Second, among members above 60 years, those that appeared to live alone (i.e. who had neither a spouse nor a dependent child in the fund), spent 14 percent more than the mean of all pensioners. Finally, a dummy for members that appear to have died in the respective year (because they were no longer in the data set in the following year) was highly significant with a coefficient of roughly 12,000 DM for members under 60 and 6,000 DM for members over 60. Altogether, these three variables raise adjusted R-squared from 0.0527 to 0.0709. Although it may be difficult to base risk adjustment payments on the percentage of singles among fund members, income is more easily observable, and a retrospective lump-sum payment for every deceased plan member, which was first proposed by Beck and Zweifel (1998), is also feasible. This would greatly reduce insurers' incentives to attract high-income persons and to fend off applicants with an elevated risk of dying.

A related issue is whether geographic variation in health care spending should be incorporated into a system of risk adjustment. If it is included, then people living in underserviced areas implicitly subsidize those living in well- or overserviced areas,

[3] See the chapter "Equity in Health and Health Care" by Fleurbaey and Schokkaert in this volume for a broader discussion of equity issues in health care.

which can be considered unfair. There are adjustments to subsidies in US Medicare for high-medical-cost areas relative to low-cost areas, but the community rating for the under 65 population envisioned under US health reform will allow substantial geographic variation in premiums that reflects geographic variation in medical costs. One of the problems here is that we do not know whether and to what extent those who suffer from higher local costs might also be getting care that is better or more technologically sophisticated.

Finally, some of the characteristics a regulator would want to include in X are subject to plan manipulation, particularly because the data for risk adjustment is usually based on insurance claims provided by the plan. When payment is based on diagnostic coding of insurance claims, plans have strong incentives to code more aggressively. Song et al. (2010) examine how the risk scores of Medicare beneficiaries change when they move to a new geographic location based on the intensity of medical care delivery in both their original and their new location. They find that moving from a low-intensity to a high-intensity practice-style area was associated with a substantial increase in the number of diagnoses coded. This suggests that not only are current methods of calculating risk scores possibly ineffective in disentangling exogenous health status from plan efficiency, but that risk scores may be relatively easily manipulated through treatment decisions and coding practices.

4.3.5.3. The Regulator's Perspective

Relatively little work has explicitly considered the objective of the regulator in risk adjustment. In the methodological literature, the objective when developing a system of risk adjustment is usually to maximize the predictive power of the underlying model of health care costs, which Glazer and McGuire (2000) refer to as the "statistical approach" to risk adjustment. These authors, in contrast, examine how the regulator can set payments to induce plans to behave in the way that the regulator desires. Using this principal–agent approach, they analyze whether a regulator with imperfect information on consumer risk can improve on the market outcome in a situation in which asymmetric information exists between consumers and health plans. They modify the Rothschild and Stiglitz setting to consider the case of managed care by allowing plans to manipulate levels of care across different types of health care services. In their model, consumers know their risk type but regulators and insurers receive only (the same) noisy signal regarding an individual's risk type. Modeling a regulated market in which the regulator requires community rating of enrollee premium contributions to achieve equity objectives and uses risk adjustment to minimize the inefficiency in coverage levels chosen by competing plans, they show that "conventional risk adjustment"—setting payments to plans equal to the expected cost of enrollees conditional on the noisy signal—improves outcomes relative to no risk adjustment. While in both cases high risks enroll in a contract which provides less care, both acute

and chronic, than is socially optimal, and low risks enroll in a plan providing more acute care and less chronic care than is socially optimal, the deviations from the socially optimal level of rationing are smaller in magnitude with conventional risk adjustment than with no risk adjustment.

The authors demonstrate, however, that the regulator can come closer to the socially optimal allocation by setting the payment rates for high risks higher than the expected cost and conversely setting the payment rates for low risks lower than the expected cost. The key contribution of this analysis is that, when a regulator has only a noisy signal of risk, which is often the case in practice, he or she can improve upon current methods of risk adjustment by adjusting payment rates to counteract selection incentives. Using a different modeling approach, Jack (2006) makes a similar point. Building on this work, Glazer and McGuire (2002) define the set of weights a regulator should choose for a given set of risk adjusters to achieve particular objectives regarding access and efficiency and provide an empirical example.

4.3.5.4. Risk Adjustment and Efficiency

Risk adjustment is a regulatory tool intended to reduce the inefficiency associated with community rating of health insurance premiums. Can a system of risk adjustment achieve a first-best efficiency when premiums are community rated? Work by Glazer and McGuire (2002) suggests that, even if a regulator has imperfect information on individual risk, he can improve on expected cost-based models by designing payments which counteract the residual incentives for risk selection. In other words, even if the regulator cannot perfectly observe relative costs, the possibility exists of designing payment systems which would eliminate the incentives on the part of plans to distort their services for the purpose of risk selection.

However, two problems remain for administered pricing on the supply side. The first, relating to productive efficiency, is whether the data available to the regulator will allow him to estimate relative costs in ways which promote efficient utilization. All systems of risk adjustment are implemented using claims data from a particular setting in which health care delivery is unlikely to be efficient. In this case, the estimates of relative costs across different risk types would reflect the inefficiency of the system upon which they are based. The second issue is whether administered pricing can achieve allocative efficiency. Premiums set in competitive markets provide a signal of the value of spending on health care relative to other goods and services and administered pricing schemes eliminate that signal. The US Medicare program, which has struggled in setting the average premium to induce the correct/efficient level of entry on the part of health plans, provides a good example of this challenge.[4] Glazer and McGuire (2002) identify this weakness when they note that their system of risk

[4] See McGuire et al. (2011) for a detailed discussion.

adjustment will induce first-best levels of rationing only when the regulator sets the budget at the appropriate level. Otherwise, the outcome is second best with overall levels of care either over- or underprovided depending on the budget set by the regulator. Systems which require enrollees to pay a contribution linked to the cost of a particular plan address this issue in part.

An additional challenge to efficiency faced by systems of community rating comes from the demand side. In most studies of adverse selection in health insurance markets and risk adjustment, demand is determined entirely by risk. Variation in preferences for health plans or style of health care delivery within a given risk type, however, introduces an intractable problem for community rating. When preferences vary within a risk type, risk-based pricing on the demand side is necessary to induce consumers to make efficient choices across different plans. Bundorf et al. (2008) show that the welfare loss due to inefficient matching of consumers to health plans in the absence of risk-rated premium contributions is quantitatively important in the context of US employer-sponsored coverage. Yet risk rating of enrollee contributions compromises the distributional objectives which motivated the regulation in the first place. Glazer and McGuire (2009) identify possible alternatives. They show that when taste can be used as the basis of taxation, as in the case of income, a simple tax can achieve the efficient allocation and preserve the equity criteria of uniform premiums by risk. But when taste is not observable and, thus, cannot be used as the basis for taxation, efficiency and equity, as defined as equal premium contributions for a given plan for differing risk types, cannot be achieved simultaneously. They propose a weaker version of equity, which they refer to as weak solidarity, and show how it can be achieved.

4.3.5.5. Recent Developments in the Practice of Risk Adjustment

A number of countries have implemented and refined systems of risk adjustment since the early 1990s. In this section we provide a brief overview of developments in the practice of risk adjustment in Europe (the Netherlands, Switzerland, and Germany) and the United States.

Germany, the Netherlands, and Switzerland introduced greater competition among sickness funds in the 1990s in an attempt to control rapidly rising health care expenditures and achieve a more efficient allocation of resources in the health care sector. To protect the desired extent of redistribution between risk classes, the competitive framework was accompanied by community rating and open enrollment. Moreover, all three countries introduced risk adjustment systems (RAS) (van de Ven et al., 2003).

In the US, the Medicare program has used a system of risk adjustment when paying private plans, which serve as a voluntary replacement for the traditional benefit, since the program was established during the early 1980s. The system, however, has been widely documented as inadequate in the sense that, due to favorable risk

selection, the average Medicare payment to private plans exceeded the amount Medicare would have paid for the beneficiary under traditional coverage.[5] As a result, a revised risk adjustment system which makes greater use of diagnostic information was phased in between 2000 and 2007. The recently implemented prescription drug benefit for Medicare beneficiaries, which relies on competing private plans to deliver publicly subsidized insurance coverage, also relies on a system of risk adjustment. And under the reforms specified by the Patient Protection and Affordable Care Act passed in 2010, risk adjustment will be applied both to private plans participating in newly established health insurance exchanges as well as plans operating outside the exchange.

4.3.5.5.1. Switzerland In Switzerland, the RAS was introduced in 1993 and the competitive framework in 1996. The RAS, using only the parameters of age and sex, was performed using the cell approach. For each age and sex cell, average expenditures were calculated for each of the 26 cantons separately and used as a basis for equalizing payments between sickness funds. It was expected that risk compositions of competing funds would assimilate over time so that the RAS could be removed after ten years (Beck et al., 2003).

The competitive framework is characterized by a basic insurance package, which is mandatory for all residents, and voluntary supplementary insurance packages. Basic insurance was provided by 86 private insurance companies in 2008, down from 118 ten years earlier. The four largest companies had a joint market share of almost 50 percent. Supplementary insurance, which is frequently offered by subsidiary companies, is provided at risk-based premiums, and applicants can be refused by an insurer. As more than 70 percent of the population has taken out supplementary insurance and often with the same insurer, the insurer possesses additional information on their client's health.

The premium for the basic insurance is uniform for all clients of a company and is calculated separately for three age groups (under 18, 18–25, and over 25 years) so as to guarantee zero profits. Discounts can be granted if the client accepts a higher deductible or participates in a managed care plan. Clients can change their insurer every six months. Citizens with low income can apply at their canton for a "premium subsidy" which covers the part of the premium that exceeds a certain percentage (usually 4 to 12) of their total income so that for this clientele the premium is essentially income dependent. Fifty percent of the costs of public hospitals are borne directly by the canton so that insurance companies only reimburse the other half.

4.3.5.5.2. Germany The German RAS was introduced in 1994 in the wake of a more competitive framework of the social health insurance (SHI) system. The main

[5] See McGuire et al. (2011) for a discussion of this literature.

target of the RAS was to reduce the wide spread of earnings-related contribution rates of sickness funds (which ranged from 8 to 16 percent) and thus to create a level playing field. Historically, many sickness funds catered only to particular industries or occupational groups so that their risk and earnings compositions differed widely. While before 1996 most blue-collar workers could not choose their sickness fund, free choice was introduced in 1996, mainly to remove the privileges of white-collar employees, not primarily to enhance efficiency through more competition (Buchner and Wasem, 2003).

The first RAS used the parameters age, sex, and disability status to group the insured in 360 cells, within each of which the average expenditures were calculated to determine the "contribution needs" of each sickness fund. As contributions are levied as a share of earnings up to some limit, the RAS also had to take into account the "financial power" of each fund, which is simply the sum of the contributable earnings of all fund members multiplied by an average contribution rate across funds. The difference between contribution needs and financial power was refunded by a central pool.

Since its introduction, the German RAS has been reformed several times. In 2002, insured persons who suffered from one of seven chronic conditions and who voluntarily enrolled in one of seven "disease management programs" were put into extra cells for which average expenditures were calculated separately. Moreover, in 2002 a high-cost pool was introduced which covered 60 percent of expenditures exceeding 20,450 euros in a given year. Finally, in 2009 a direct morbidity component was introduced (see below).

The German health care system is a two-tier system which consists of a social and a private tier. Since 2009 there has been mandatory insurance for every resident. Employees with earnings below a certain threshold (currently around 50,000 euros per year) are members of SHI whereas high earners, civil servants, and the self-employed can opt out of SHI and take out private insurance. While in the private tier premiums are risk rated, the SHI tier is characterized by a legally determined benefit package, free enrollment and earnings-related contributions at a uniform contribution rate, which is fixed by government, but sickness funds are allowed to levy an additional contribution, allowing for a small element of price competition. Moreover, since 2007 sickness funds are allowed to offer rebates for contracts with higher co-payments or restricted physician access. The number of sickness funds has fallen dramatically since the introduction of free choice of funds in 1996, from 642 to less than 200 in 2010. Earnings-related contributions and government subsidies (presently almost 10 percent of total SHI expenditures) go to a central fund, which distributes the money to the sickness funds according to their risk profile.

The latter is measured using the traditional risk adjusters of age, sex, and disability status, but each sickness fund receives a supplement for every client who suffers one

of 80 specific chronic diseases. Patients are identified using "diagnostic information" and drug prescriptions and allocated to one of 152 risk groups for which expected costs for the following year are determined using a linear regression.

4.3.5.5.3. Netherlands The introduction of the RAS in the Netherlands in 1991 was an important part of market-oriented health care reform which was proposed 1987 by the Dekker Committee (Helderman et al., 2005) and inspired by Enthoven (1988). Its guiding principle was that of regulated competition among health insurers and health care providers and more cost responsibility for insurers. Many of these ideas surfaced again in the health care reform of 1996. At first, the RAS used only age and sex as risk adjusters and was accompanied by a large cost reimbursement scheme containing a high-cost pool (Lamers et al., 2003; van de Ven and Schut, 2008). In 1995 the additional risk adjusters "urbanization" and "income" were included, but in 2000 "previous expenditures" were used again, turning the RAS into a partial cost reimbursement system. In 2002 "previous expenditures" were replaced by "pharmaceutical cost groups" and in 2004 "diagnostic cost group" and "employment status" (Douven, 2007).

Since the health care reform of 2006, there has been a mandatory basic insurance in the Netherlands, which is provided by five large private insurance companies in a strictly regulated environment with annual free enrollment, a regulated benefit package and community rating. Insurers are allowed to selectively contract with health care providers and to run their own pharmacies (van de Ven and Schut, 2008). Fifty percent of the revenues of basic insurance plans are financed through income-related contributions, which flow into a central RAS fund (Douven, 2007). The remaining 50 percent are financed through a per-capita premium which is levied directly by the insurance company from all insured persons above 18 years of age. The premium is not risk related but differs between insurers. Insurers may give a rebate of up to 10 percent for group insurance contracts (57 percent of all insureds) and further deductions for co-payments exceeding 150 euros per year (van de Ven and Schut, 2008). About two-thirds of the population receive a premium subsidy from the government, which depends on the average premium so that incentives to choose a cheap contract remain intact.

The revenues from the income-related part flow from the central fund to the insurers in the form of risk-related allocations (Douven, 2007). The procedure consists of several steps (van de Ven and Schut, 2008). In the first step, the expenditures of the coming year are estimated to determine the required contribution rate to finance one-half of total expenditures. In the second step, individual expenditures are estimated, which determine individual risk-related transfer payments. Finally, switches between insurers are estimated to calculate allocations to insurers (Douven, 2007). Individual health expenditures are estimated on the basis of age, sex, urbanization,

pharmaceutical cost group (20 groups), and diagnostic cost group. Since 2008, "socio-economic status" has been used as an additional risk adjuster to equalize previously unexplained differences between insured persons who were previously members of social as opposed to private health insurers. Income was not suited to explain this difference (Douven, 2007). Although the Dutch RAS has many risk adjusters, calculated budgets and true expenditures can differ widely, and at the end of each year this difference is partly covered in an attempt to further reduce incentives for risk selection. This is achieved via retrospective partial cost reimbursement and a mandatory high-cost pool which covers 90 percent of all costs exceeding 20,000 euros (van de Ven and Schut, 2008).

4.3.5.5.4. The United States The US Medicare program provides publicly funded health insurance for aged and disabled populations. Since the early 1980s, Medicare beneficiaries have had the option of replacing this traditional coverage with insurance from a private plan.[6] Participating plans sign an annual contract with the Centers for Medicaid and Medicare Services (CMS) agreeing to provide benefits to beneficiaries. Plans must provide a minimum level of benefits equivalent to that of traditional Medicare, may, and often do, provide additional benefits, such as lower cost-sharing or additional covered services, and may charge beneficiaries a premium for enrolling. The beneficiary premium must be community rated within a given service area and plans must accept all beneficiaries who would like to enroll. Plans are allowed to use managed care techniques to coordinate care.

For beneficiaries who choose to enroll in a private plan, the Medicare program makes a risk-adjusted capitated payment to the health plan on the beneficiary's behalf.[7] Historically, CMS linked plan payments to the level of Medicare spending among FFS enrollees in a geographic area and adjusted the payment for demographic characteristics of enrollees. Payments were set at 95 percent of the adjusted average per capita cost (AAPCC) in the county, calculated separately for aged and disabled populations, and the individual characteristics used for risk adjustment were age, sex, Medicaid enrollment, institutional status (nursing home resident), and working status. This "demographic model," however, explained only 1 percent of the variation in Medicare spending, creating considerable opportunity for favorable risk selection into private plans. Correspondingly, several studies found that, due to favorable selection, the Medicare program paid private plans more for enrollees than it would have cost if they had remained in the traditional program.[8]

[6] Traditional Medicare coverage is composed of Parts A and B, which correspond roughly to inpatient and outpatient services, respectively. A private plan, technically Medicare Part C, replaces both Part A and Part B coverage. Part D, Medicare coverage for prescription drugs, was enacted in 2003 and first became available in 2006. Prior to the enactment of Part D, Medicare did not cover outpatient prescription drugs.

[7] See Pope et al. (2004) for a detailed description of the development of this system.

[8] See McGuire et al. (2011) for a discussion of the effect of Medicare Part C payment methods on program costs.

In response to these concerns, the Balanced Budget Act of 1997 mandated the eventual risk adjustment of plan payments based on enrollee health status.[9] In 2000, CMS began risk adjusting payments using information on inpatient diagnoses. Initially, the system was limited to inpatient diagnoses due to the lack of reliably coded data from the outpatient setting, and only 10 percent of the weight was based on the modified risk adjustment system due to the concern that it created incentives for plans to hospitalize patients in order to increase payments. Subsequent legislation mandated that, by 2004, CMS base its risk adjustment on data from both inpatient and ambulatory settings, and, in 2004, CMS introduced a more comprehensive system of risk adjustment based on hierarchical condition categories (HCCs) derived from both inpatient and outpatient data. Risk-adjusted payments represented 30, 50, and 75 percent of a blended rate in 2004, 2005, and 2006, respectively. Starting in 2007, risk adjustment was based entirely on the HCC model.

The new model, which added indicators of diagnoses to the existing demographic model, uses diagnoses from a base year to predict utilization in a subsequent year. Indicators of diagnoses were developed by mapping over 15,000 ICD-9 codes to approximately 180 condition categories, from which only a subset, selected based on their predictive power and other criteria, was ultimately used in the model. Among the clinically similar condition categories included in the model, a hierarchy is imposed so that only the most severe condition within a group is coded. Individuals may be coded with multiple conditions when the conditions are not clinically similar. Ultimately the model included 70 diagnostic categories, six diagnostic category interactions, and five interactions of diagnostic categories with indicators of entitlement by disability. The new model, which was calibrated using 1999–2000 Medicare claims data, has greater predictive power than the demographic model, explaining about 10 percent of variation in costs among beneficiaries enrolled in traditional coverage.

Risk adjustment is also a key component of the Medicare Part D Prescription Drug Benefit. The Medicare program extended outpatient prescription drug coverage to beneficiaries beginning in 2006, and the new benefit relies exclusively on competing private insurance plans to deliver publicly subsidized coverage.[10] Private insurance plans contract with Medicare to provide coverage in specific regions. While the program has defined a standard benefit, plans may deviate from that standard in designing coverage providing they meet certain standards with respect to actuarial equivalence, and beneficiaries must pay the full incremental cost for a particular plan relative to a standard plan when they choose to enroll in a particular plan. Thus, plans compete for enrollees based on benefit design, price, and service. Overall Medicare subsidizes

[9] The BBA of 1997 also marked the beginning of a series of payment reforms which increased the level of payments to private plans irrespective of risk and these changes also significantly increased Medicare spending. See McGuire et al. (2011) for a comprehensive discussion of these changes.

[10] See Duggan et al. (2008) for a detailed description of this program.

the premium by about 75 percent and provides additional subsidies for low-income beneficiaries.

Medicare's payments to plans and enrollee contributions are determined through a competitive bidding process. Each year participating plans submit bids to the Medicare program for covering a beneficiary of average health. The enrollee premium for a plan is the difference between the plan bid and 74.5 percent of the national average bid, and the enrollee premium does not vary by individual risk.[11] The base payment to the plan, which is the difference between the plan bid and the enrollee contribution, is adjusted for enrollee risk.

The general approach to risk adjustment is similar between the Part C and Part D benefits.[12] The system is based on a model which uses individual diagnoses derived from insurance claims in one year to predict utilization in a subsequent year. Because data on prescription drug utilization for Medicare beneficiaries were not available prior to the implementation of the program, the model is based on prescription drug expenditure data from Federal retirees and Medicaid beneficiaries which were then linked to Medicare inpatient and outpatient utilization data. As in the case of risk adjustment for private managed care plans, indicators of diagnoses were generated by mapping ICD-9 codes into clinical conditions and then determining which clinical conditions were most predictive of subsequent year utilization. Both the aggregation of ICD-9 codes into clinical conditions and the relationship between clinical conditions and utilization differed between the two settings. The final prescription drug spending model used for risk adjustment includes indicators of age and sex, an indicator of disability status, 84 RX-HCCs (disease indicators), and the interaction of disabled status with three different disease indicators. As in the case of the HCC model, coding for indicators for a subset of clinically similar conditions is hierarchical based on disease severity while the remainder of condition indicators are additive. The risk-adjusted payment is adjusted by a scaling factor for beneficiaries receiving low-income subsidies and those who are institutionalized, reflecting the higher expected spending for these groups.

Brown et al. (2011), analyzing the transition from the demographic to the HCC risk adjustment system, provide the first evidence of the effects of risk adjustment in the US Medicare program. They find that risk selection along the dimensions measured in the HCC model declined as risk adjustment was phased in, suggesting that more detailed risk adjustment did create incentives for plans to enroll higher-risk beneficiaries for whom they received greater compensation. However, plans also responded to incentives for greater risk selection along the unmeasured dimensions. In other words, plans increased the degree of risk selection conditional upon the risk scores. Their analysis suggests that these responses ultimately increased the cost to the

[11] An adjustment also exists for individual reinsurance subsidies received by the plan.

[12] See Robst et al. (2007) for a detailed description of the development of the system of risk adjustment for Medicare Part D.

government of the Medicare program, by increasing the difference between the amount Medicare paid private plans for enrollees and the amount they would have paid for these beneficiaries in the traditional program.

5. EMPIRICAL EVIDENCE ON INSURANCE MARKET PERFORMANCE

5.1. Risk Variation and the Demand for Insurance

How should the demand for insurance vary with the person's risk or perception of risk? This simple question actually has a complex answer. One's intuition is that insurance is more valuable to a person with larger expected benefits from that insurance, and so higher risks should be both more likely to demand a given policy (compared to none) and more likely to demand more generous coverage than lower risks. But in competitive insurance markets new high-risk buyers will face higher premiums (both in total and for increments in coverage) than low-risk buyers. Which should win out?

The most transparent answer, given by Ehrlich and Becker (1972), is that, absent income effects on the demand for insurance, there should be a tie, so that the demand for insurance is independent of risk levels in fully informed competitive markets. They conclude that demand should depend only on the loading factor, and not on the probability or amount of loss.

As with all comparative statics results in consumer choice, these statements about demand can be changed if there are large income effects. Compared to being a low risk, someone who is a high risk will pay a high premium (or without insurance face higher expected expense) which reduces real income. That reduction in real income may then affect the demand for insurance if risk aversion (absolute or relative) varies with income or wealth. That is, whether higher risk affects the demand for insurance depends on how risk aversion varies with (expected) real income. The demand for insurance could then either rise or fall as risk and premiums change: if high insurance premiums take a larger bite out of my income because I am a high risk, I might increase demand because I am poorer and more vulnerable to losses or reduce my demand because insurance loading and services are luxuries I can no longer afford.

What about the common concern that risk rating will make health insurance "unaffordable" for low-to-moderate income high risks? This can happen if income positively affects the demand for medical care but insurance pricing does not take income into account. If I am a low-income low-risk person but am faced with premiums that vary only with risk and not income, I will be paying a premium that largely depends on the choices of higher-income people—a premium I may regard as excessive relative to the (smaller) expected out-of-pocket payment I would face without insurance. If income effects on the demand for care are large, insurance markets

will need to be segmented by income to function properly. Bundorf et al. (2010) find empirical evidence consistent with these types of income effects in the employment-based large group market in the US. Rather than observing low rates of coverage among low-income high risks, however, they observe relatively low rates of coverage among low-income low risks, which they attribute to the effect of pooled premiums on the affordability of coverage among low risks. In other words, among low risks, pooled premiums represent a larger barrier to coverage among those with low rather than those with high incomes.

The effect of risk on demand for coverage may also be affected by the extent to which preferences for coverage, due to factors such as risk aversion, are correlated with risk. For example, if high risks are also more risk averse, then they will have stronger demand for coverage, even when the load does not vary by risk type.

5.2. Empirical Evidence on the Functioning of Insurance Markets

Little evidence exists on the extent to which adverse selection affects the operation of unregulated, voluntary individual health insurance markets. Because health insurance markets are highly regulated in nearly all countries and the purchase of coverage is compulsory in most, few opportunities exist to observe consumer and insurer behavior in a truly competitive market. Even when public programs do not explicitly replace private markets for the majority of the population, institutions have evolved which differentiate the market from the textbook single-period risk-rating model. While the purchase of coverage is voluntary in the US, employer-sponsored health insurance is highly subsidized through the exclusion of premiums from taxable compensation and the bulk of the population obtains group insurance through employers. Because employers control most aspects of benefit design, including insurer entry and enrollee premium contributions, the performance of the employer-sponsored market likely differs from that of a competitive individual market. In addition, both federal and state regulation influences the employer-sponsored market in a variety of ways. Although insurance markets in the US are generally regulated at the state level, employers who self-insure are exempt from state regulation and regulated instead by a series of federal regulations such as (but not limited to) non-discrimination requirements, restrictions on the types of factors that may be considered in setting employee premium contributions, and limitations on the exclusion of pre-existing conditions. State regulation, which applies primarily to the individual and small group markets in the US, affects many aspects of insurer behavior including mandated benefit requirements and restrictions on insurer rating practices. States have adopted very different regulatory models.

The US individual health insurance market, particularly in states with relatively weak regulatory environments, is likely the context which most closely resembles an

unregulated market. Yet few studies in this context exist. The individual market is relatively small and self-selected in the US—most people with private health insurance obtain it through an employer, and even more surely could if they were willing to take a job at a firm that offered coverage (such as Starbucks)—and obtaining data on both consumer and insurer behavior in this context is challenging. In addition, the existence of the tax-favored employer-sponsored market likely influences the functioning of the individual market, making it virtually impossible to disentangle the two. As a result, most studies of adverse selection in insurance markets are based on either highly regulated markets for individual primary or supplemental coverage or on employment-based group markets (often with tax subsidies but with relatively low explicit regulation).

While many other countries rely on private health insurance to varying degrees, the markets vary substantially across as well as within countries. In the Netherlands, for example, the vast majority of the population purchases primary coverage from private insurers who compete in a highly regulated, "managed competition" setting. In Germany, in contrast, the bulk of the population purchases primary coverage through a regulated social insurance system, but civil servants and high-income individuals are allowed to opt out and purchase coverage in a private insurance market.

The degree of heterogeneity in the institutional environments in which studies of risk selection take place suggests caution in generalizing the results from a particular study to an alternative setting. In effect, any empirical setting to be examined involves an imperfectly understood mix of employer, insurer, and regulatory features that affect risk variation, and so cannot be extended to some other setting involving equally mixed policies.

5.3. Empirical Methods Used to Examine Risk Selection

The key empirical challenge in identifying risk selection in markets for health insurance, no matter what the institutional arrangement, is separating the effect of risk on the purchase of coverage from the effect of insurance on the utilization of medical care. Economic theory predicts both that high risks will purchase greater quantities of insurance than low risks at a given premium (adverse selection) and that, holding *ex-ante* risk constant, insurance coverage will increase the probability and/or size of a loss (moral hazard) (Pauly, 1968; Rothschild and Stiglitz, 1976). Because both theories predict that *ex-post* expenditures are positively correlated with the amount of insurance, it is difficult to disentangle empirically adverse selection from moral hazard when comparing medical care expenditures among people with differing levels of insurance coverage (Chiappori, 2000).

Researchers have used a variety of approaches to address this issue. "Switcher studies" compare, among people enrolled in a particular type of coverage in a given year,

the utilization of people who changed their coverage with those who did not in the subsequent year. The advantage of this type of study is that, by limiting the study sample to people enrolled in the same type of coverage in the initial year, comparisons of expenditures are not biased by differences in the effects of plan characteristics on the utilization of health care. The primary weakness is that utilization one period prior to enrollment may not necessarily be representative of longer-term differences between the two groups. If pre-enrollment differences in utilization represent transitory shocks, rather than permanent differences in health status and medical care utilization, the average difference between enrollees with different types of coverage will be smaller than the marginal difference (Hellinger, 1987; Brown et al., 1993). If, as seems to be the case, prior period spending is one thing that motivates changing insurance, the identification problem is even worse.

Other studies compare indicators of health status, rather than measures of expenditures, between two groups. By examining all enrollees, not just those who switch plans, these studies provide estimates of the difference in health status of the average, rather than the marginal enrollee. However, in order to avoid bias created by the effect of plan characteristics on utilization, rather than examine medical care expenditures or utilization, researchers usually examine measures of health that are less likely to be influenced by plan characteristics, such as indicators of health conditions and self-reported health status. Most studies examine multiple measures and find evidence of favorable selection along some, but not others, and few translate the findings into differences in expenditures, making it difficult to determine whether the observed differences in health represent substantive differences in health expenditures. The potential endogeneity of measures of health represents an additional concern. To the extent that insurance coverage increases the rate of diagnosis of conditions (in the case of indicators of chronic conditions) or the effectiveness of treatment (in the case of self-reported health status), these studies will overestimate the extent of risk-based selection.

As will be shown, studies using these methods have been concerned primarily with documenting simply whether coverage varies with risk—whether higher-risk people are more likely to have health insurance or to have more generous coverage than lower-risk people. Fewer studies provide evidence on whether risk-based selection is indeed due to asymmetric information between insurers and consumers. More precisely, to attribute risk selection to asymmetric information, it is necessary to demonstrate that those who have insurance and have more generous insurance are higher risk conditional on the information available to and used by insurers. As mentioned earlier, potential alternative explanations for a positive relationship between risk and coverage include moral hazard, income effects, and a correlation between risk and preferences for insurance. Finkelstein and McGarry (2006) provide an example of the role of preferences in the market for long-term care insurance. In this case, however, the absence of a difference in claims rates between those with and without coverage is

due to a negative correlation between risk and preferences for coverage. Thus, documenting a positive correlation between risk and coverage is neither necessary nor sufficient to demonstrate the existence of risk-based selection due to asymmetric information.

This observation that, in many markets, insurers are either prevented by regulation or voluntarily forego using information that is correlated with claims motivates the "unused observables" test for adverse selection proposed by Finkelstein et al. (2006). A correlation between these unpriced observable risk factors and the purchase of health insurance, conditioning on the factors used in pricing, demonstrates the existence of adverse selection. In the case of health insurance markets, regulation often prevents insurers from using easily available information, such as age, sex or the presence of health conditions, in pricing coverage. Thus, this approach implicitly underlies analyses of the effects of these types of regulations on rates of coverage which tend to find that rating restrictions increase rates of coverage among high risks and reduce rates of coverage among low risks (Buchmueller and DiNardo, 2002; Monheit and Steinberg Schone, 2004; Simon, 2005; Davidoff et al., 2005; Bundorf and Simon, 2006; Pauly and Herring, 2007; Lo Sasso and Lurie, 2009).

Perhaps due to the availability of richer data sets and the development of more sophisticated empirical methods, a growing literature is examining risk selection due to asymmetric information in health insurance markets using structural models (Cardon and Hendel, 2001; Bundorf et al., 2008; Carlin and Town, 2010; Bajari et al., 2010). A study by Cardon and Hendel (2001), based on nationally representative survey data, was the first in this line of research. Cardon and Hendel use variation across workers in the set of plans offered by their employer to identify the effect of plan type on health care utilization (moral hazard) among similar consumers (based on observable characteristics).[13] They then test whether *ex-post* differences in expenditures are related to insurance choice after accounting for observable characteristics. This approach not only distinguishes between moral hazard and adverse selection in accounting for utilization differences among people enrolled in different types of coverage, but also allows for private information about risk type to influence choice. The authors find little evidence of selection on unobservables, suggesting that adverse selection may not be an important factor in unregulated group health insurance markets because the individual characteristics which affect choice could easily be incorporated into price. Its findings are not necessarily consistent with the absence of adverse selection in the market for employer-sponsored coverage. This is because selection on observables may represent adverse selection if the observable characteristics are not used in pricing. In other words, if neither employee premium contributions nor wage

[13] The assumption that choice sets are exogenous with respect to worker demand for health insurance is controversial. Empirical evidence suggests that the choice sets offered by employers are related to employee preferences for health insurance (Moran et al., 2001; Bundorf, 2002).

offsets vary with observable characteristics of consumers that influence plan choice, then selection on these characteristics represents selection on unpriced risk factors. This distinction demonstrates the importance of carefully considering the institutional features of a particular market when interpreting evidence on the extent of risk-based selection from a particular setting. Einav et al. (2007) propose an alternative method for identifying adverse selection in health insurance markets. Their method requires a situation in which the premium for insurance varies exogenously and the analyst has information on consumer demand as well as insurer costs. They demonstrate how, under these conditions, the expected benefits cost curve of the insurer will respond endogenously to price in the presence of risk selection. When adverse selection is present, insurer cost increases with price, and when advantageous selection is present, insurer cost declines with price. The intuition is that at high premiums only high risks will purchase, and will then incur high benefits costs, whereas at lower premiums lower risks will buy. Of course, the firm will only be able to operate at those points on this "cost" curve at which demand is adequate to cover these costs. Pauly (1974) and Pauly and Zeng (2004) provide simple examples of these kinds of models.

A key advantage of these more structural approaches to analyzing risk selection in health insurance markets is that, by estimating models of both demand and costs, the analyst can evaluate the welfare impact of adverse selection (Einav et al., 2010b). It is necessary, however, to identify the relevant benchmark when interpreting these estimates. For example, Carlin and Town (2010), analyzing the employer-sponsored market, compare the observed and potential alternative uniform contribution policies relative to the most efficient uniform contribution. In other words, this study takes as its benchmark an environment with limited individual variation in premiums and identifies the welfare loss of alternative policies relative to this second-best, effectively measuring the degree of welfare loss associated with the policies the employer did adopt relative to what the employer could have adopted within the existing institutional constraints. Bundorf et al. (2008), in contrast, calculate the welfare costs of different pricing policies relative to an alternative benchmark—an individual's single-period risk-rated premium. While use of this benchmark allows them to consider how existing institutional arrangements influence the efficiency of health insurance markets, the benchmark does not represent first-best. In this case, however, risk-rated premiums expose consumers to dynamic reclassification risk, and for the purpose of policy analysis, one would also want to consider this source of inefficiency. Finally, estimates from Einav et al. (2010a), using data from the employer-sponsored market, compare a hypothetical competitive uniform contribution relative to the most efficient (second-best) uniform contribution, providing little insight into the extent of welfare loss in the current employer-sponsored market. This work is related to the problem of optimal premium differentials in group insurance discussed in section 3.2.2.

More generally, the absence of a benchmark unregulated insurance model is an important limitation of these types of calculations. Interestingly, "out-of-equilibrium" pricing enables researchers to make welfare calculations by allowing them to observe how different types of consumers behave when facing different prices, which are unrelated to demand. However, the fact that "out-of-equilibrium" is widespread in health insurance markets indicates that it is virtually impossible to know what the relevant competitive market outcome would look like.

5.4. Empirical Evidence on Risk Selection

In this section, we organize evidence on risk selection by whether the coverage is primary or supplementary and whether the study examines the consumer's decision to purchase any health insurance or what type of coverage.

5.4.1. Primary Coverage

5.4.1.1. Risk Selection into Coverage

Only a few studies have examined the role of risk in whether people purchase health insurance in the individual market in the US. While Pauly and Herring (1999) did not find a statistically significant relationship between risk (measured by expected medical expenses) and insurance coverage in the individual market, in more recent work, they find that high-risk people, based on the presence of chronic conditions, are *less* likely than lower risks to purchase coverage, particularly in states without regulations restricting the rating practices of insurers (Pauly and Herring, 2007), which could be driven by income constraints.

Studies comparing measures of health status of uninsured and insured people in the employer-sponsored market often find that the insured are more likely to report having a chronic condition but also are less likely to report poor health (Monheit and Vistnes, 1994; Bernard and Selden, 2002). Studies examining people offered health insurance from an employer have found that those who decline coverage and are uninsured are healthier than those who enroll on some physical health measures but less healthy on others (Blumberg et al., 2001). Similarly, those who decline coverage are more likely to report poor health, yet less likely to have a high-cost chronic condition than those who enroll (Bernard and Selden, 2002).

Pauly and Herring (1999) did not find a statistically significant relationship between risk (measured by expected medical expenses) and insurance coverage in the large group market, although they did find that high-risk, lower-income people working for small firms were less likely to obtain insurance than otherwise similar low-risk people. Bundorf et al. (2010), in contrast, find a positive correlation between health risk, as measured by expected expenditures based on a more detailed list of self-reported health conditions, and coverage in the employer-sponsored market. The existence of a positive relationship between risk and coverage suggests that the

premiums consumers face for coverage in this setting, in the form of employee contributions and potential wage offsets, incorporate less information about individual risk than consumers use when deciding whether to obtain coverage. In the large group market, the positive relationship between health risk and coverage is stronger for people with low than for people with high incomes. The finding of a moderate degree of adverse selection into employer-sponsored coverage is consistent with evidence from Bhattacharya and Vogt (2006) that greater job-specific capital among workers is associated with higher rates of employment-based coverage and that this effect is strongest among low-risk workers. The results from Bundorf et al. (2010) are consistent with this type of separation, particularly among low- and medium-income workers. Switching costs may be adequately high in large firms, particularly for high-income workers, to promote pooling among this group.

5.4.1.2. Risk Selection among Health Plans

A relatively large literature examines risk selection among health plans. Van Vliet (2006) analyzed data from the Dutch sickness fund system in the first eight years after the introduction of free plan choice (1994–2002). Of a total of 10 million sickness fund members, 130,000 switched plans. Those who switched plans incurred expenditures that were some 40 percent lower than the average insured so that a uniform capitation payment from the central fund to the insurer in the year 2002 would overpay for switchers (of the years 2000 and 2001) by 50.2 percent. However, switchers were on average younger and thus, when adjusted for age and sex composition, this overpayment drops to 9.3 percent. Adding region and social entitlement brings this number down to 5.7 percent and finally, adding the medical parameters, to an insignificant 1.4 percent. The author concludes that although switchers are on average better risks, the Dutch RAS is able to eliminate the difference, in particular since the addition of medical information in the risk adjustment formula.

In the US, literature examining risk selection among plans has focused primarily on the decision to enrol in an HMO or other type of managed care plan rather than other less restrictive forms of coverage.

In the case of the US Medicare program, beneficiaries are able to replace their traditional publicly managed benefit with a private alternative, and a large literature examines the relationship between beneficiary risk and enrollment in private plan alternatives, with most studies indicating that private plans experience favorable selection (Hellinger, 1995; Hellinger and Wong, 2000; and Mello et al., 2003 provide reviews).

Studies have also examined risk selection among different types of plans in the US employer-sponsored group market. While early studies generally found evidence of favorable selection into HMOs relative to indemnity plans (Hellinger, 1995), studies from the 1990s and later of the extent to which managed care plans (generally

HMOs) experience favorable selection relative to less managed plans (generally PPOs or indemnity plans) produce mixed evidence (Hellinger and Wong, 2000). In general, switcher studies provide the strongest evidence of favorable selection into HMOs relative to less tightly managed plans. Nicholson et al. (2004) find evidence of favorable selection into HMOs based on both demographics and other measures of health status such as the existence of chronic conditions. One switcher study provides evidence of favorable selection based on demographic characteristics into HMOs relative to an indemnity plan (but not relative to a PPO) (Altman et al., 1998). Another finds evidence of unfavorable selection for HMOs in switching behavior for maternity admission but favorable selection for non-maternity admissions (Robinson et al., 1993).

Studies examining differences in the average characteristics of enrollees in different types of plans in the US employer-sponsored market generally do not find evidence of large differences in risk. Some studies find little to no evidence of differential selection based on either demographics or health status (Fama et al., 1995; Goldman, 1995; Schaefer and Reschovsky, 2002; Polsky and Nicholson, 2004). Others examine only age and find no evidence of favorable selection (Florence and Thorpe, 2003). Other studies of settings in which multiple managed care plans are offered show that some experience favorable selection and others do not (Robinson and Gardner, 1995; Shewry et al., 1996). A series of studies conducted in a large employer group find evidence of favorable selection into an HMO relative to an indemnity plan based on average demographic and health characteristics of enrollees but suggest that the characteristics of PPO enrollees are more similar to those of HMO enrollees (Cutler and Zeckhauser, 1998; Altman et al., 2003). In a setting in which employers offer both tightly managed HMO products and more loosely managed PPOs, Bundorf et al. (2008) find that, while no single plan consistently experiences unfavorable selection, different plans experience unfavorable selection along differing components of risk. Similarly, while Carlin and Town (2010) find that an HMO experiences favorable selection along measures of both demographic characteristics and health status relative to a high-deductible plan, a different managed care plan experiences favorable selection based on age but unfavorable selection based on health status. Bajari et al. (2010) find evidence of adverse selection into a firm's PPO plans relative to its HMO.

Bundorf et al. (2008) suggest that the absence of systematic risk selection by plan type in more recent years may be due to the evolution of the types of insurance products in the market. In particular, over the last 20 years, the market for health insurance has witnessed a dramatic shift from traditional insurance products which differed primarily based on their cost sharing to managed care products which adopt a combination of supply- and demand-side utilization controls. Because people are likely to vary in their preferences for different types of utilization controls, this shift from vertical to horizontal product differentiation may have increased the relative importance of sorting based on preferences relative to sorting based on risk. Similarly, others have

proposed that individual characteristics such as income (Schaefer and Reschovsky, 2002) or existing relationships with providers are more important determinants of plan choice (Goldman, 1995). Finally, employers may either intentionally or unintentionally set premium contributions in ways that reduce the extent of risk selection by generally subsidizing more expensive plans (Pauly and Herring, 2001).

Most studies in the emerging literature quantifying the welfare impact of adverse selection are based on the employer-sponsored market. Carlin and Town (2010) analyze the welfare implications of alternative employee contribution policies relative to an employer's observed policy, restricting their analysis to those that vary across plans but not across workers. They estimate that observed contributions lower welfare by about $13 per employee per year relative to the optimal uniform contribution. Although employee contributions under the optimal uniform contribution policy differ dramatically from observed contributions for many of the plans, the resulting welfare loss is relatively small because employees' demand is extremely price inelastic. Thus, a change in plan premiums due to a change in the risk composition of enrollees has little effect on choice behavior. Consistent with these findings, Handel (2010) finds that, while switching costs prevent consumers from adapting their decisions appropriately when the market environment changes, switching costs also reduce the degree of welfare loss due to adverse selection. Bundorf et al. (2008) also find relatively little welfare loss associated with a firm's actual choice of a uniform contribution relative to the most efficient uniform contribution. However, they estimate that individualized pricing—varying contributions based on individual risk—generates a larger increase in welfare.

5.4.2. Private Insurance Supplementing Public Coverage

Many studies of the interaction between private insurance and publicly funded programs examine the experience of the US Medicare program for the aged and disabled. US Medicare beneficiaries may choose to supplement their public coverage with a private plan purchased in a highly regulated market. For example, beginning in 1992, federal regulation restricted the sale of insurance supplementing Medicare to a set of ten predefined plans and required guaranteed issue and prohibited the use of health information in rate setting for 65-year-olds newly entering the program. These regulations have been adjusted over time to accommodate changes in the market such as the diffusion of PPO products and the implementation of the Medicare prescription drug benefit. In addition, particular states have alternative forms of benefits standardization and rating restrictions which supersede the federal requirements.

Studies examining the correlation between beneficiary characteristics and the likelihood of purchasing supplemental insurance produce mixed evidence on the direction and importance of the effect of health. Some studies find that better health is positively associated with the purchase of supplemental health insurance (Del Bene and Vaughan, 1992; Short and Vistnes, 1992; Vistnes and Banthin, 1997–1998), others

find no relationship (Browne and Doerpinghaus, 1994–1995), and others find evidence of unfavorable selection into these plans (Wolfe and Goddeeris, 1991). Studies that examine the effects of chronic conditions often find that some conditions, but not others, affect rates of coverage (Ettner, 1997; Vistnes and Banthin, 1997–1998). One study finds that the effect of health status on the likelihood of purchasing Medigap coverage depends on beneficiary knowledge about Medicare benefits (Davidson et al., 1992).

More recent work finds that, while those with higher expected medical expenditures due to the existence of health conditions are more likely to purchase supplemental coverage, this effect is offset by the effects of other characteristics, such as income and cognitive ability, that are both positively correlated with the purchase of supplemental coverage and negatively correlated with medical care spending (Fang et al., 2008). The net effect is advantageous selection into supplemental plans. Not only does this potentially provide insight into the conflicting findings of earlier work, but it also has interesting implications for the form of the welfare loss due to risk selection in this market. If premiums do not vary with individual risk, advantageous selection will tend to lead to overinsurance among high risks in this market. Empirical evidence on adverse selection in markets for private supplemental insurance is less ambiguous in the case of prescription drug insurance. National legislation implemented in 1992 restricted the sale of private insurance supplementing US Medicare to a set of ten predefined plans, with prescription drug benefits available in only three of those plans. Anecdotal evidence pointed to severe problems of adverse selection for the prescription drug coverage. Medigap policies that provided drug coverage typically imposed an upper limit on benefits—usually $1,000 to $1,200 per year, but the difference in premiums between Medigap plans with drug coverage and those without found that the premium differential was almost as large as the coverage limit, surely evidence of severe adverse selection. Relatively few beneficiaries enrolled in these plans (US GAO, 2002) and the incremental premium for these benefits relative to the average actuarial value of the benefits was relatively high (Robst, 2006).

In 2006, highly subsidized prescription drug coverage was added to the US Medicare benefit. The new benefit is provided exclusively by private insurers who compete in a highly regulated setting. Beneficiary premiums are community rated and payments to plans are risk adjusted. While insurers are required to offer coverage at least as generous as a specified minimum, they are allowed to deviate from the standard benefit in two ways. First, they are allowed to offer modified plans that are actuarially equivalent to the standard benefit. Second, they may offer coverage that is more generous than the standard benefit. The costs of enhanced benefits, however, are not subsidized by the government and are not subject to risk equalization. Because the program is relatively new, few studies have examined the extent of risk selection in this market.

A smaller literature exists on risk selection in countries other than the US, focusing primarily on supplemental coverage. Doiron et al. (2008) find for Australia that persons with chronic health conditions are more likely to purchase supplemental private health insurance, which provides access to private hospitals, which would support the adverse selection hypothesis. They propose that the well-documented positive relationship between self-reported health status and coverage is driven by a positive correlation between self-reported health status and other characteristics, such as income and risk aversion, associated with greater demand for health insurance. Shmueli (2001) finds for supplemental coverage in Israel that sicker individuals are more likely to apply for private health insurance, but are also more likely to be rejected.

6. INTERACTIONS BETWEEN RISK VARIATION AND OTHER TOPICS IN HEALTH ECONOMICS

6.1. Risk Variation and Insurance Pricing to Consumers: Bonus-Malus Systems, Discounts for Good Health Practices

Well-insured people will have higher levels of health spending, other things being equal, than less well-insured or uninsured people either under moral hazard or adverse selection. Indeed, one of the main empirical challenges in measuring one influence is to separate out the other. But is there additional interaction between risk variation and moral hazard? There might be correlation. Adverse selection occurs in the standard Rothschild–Stiglitz model by offering insurance with higher levels (in health insurance terminology) of cost sharing, but that cost sharing in turn affects expected expense if moral hazard is present. If two services had the same distribution of risk in the no-insurance world but different degrees of moral hazard, the one with the higher moral hazard would optimally have a higher level of coinsurance—but then the equilibrium in an adverse selection situation would be different. Whether these considerations are important for health insurance in general is not known, but it is possible that the market for insurance for outpatient mental health care—which is known to display high-demand elasticity—is more troubled by adverse selection (at least under premium regulation) for this reason.

A more serious possible interaction arises when insurance premium changes are tied to the experience of the insured, as under bonus-malus schemes. Then, as noted by Chiappori et al. (2006), there can be interaction between risk levels and moral hazard. The potential change in next year's premium if I use more care acts to control moral hazard. But, depending on how experience rating works, if I make no claims because I did not get sick that period (and not because I was a low risk), the resulting low premium may encourage moral hazard (by encouraging more generous coverage

in the next period). The key issue here is whether the "bonus" is calibrated properly to risk, or just to spending levels.

A third interaction between risk variation and other aspects of the health care system arises when providers are paid on some basis other than fee for service. Higher than average risk then means lower profits to a provider paid capitation or bundled payment, and lower utility to a provider paid a salary. Providers may react by trying to attract lower risks, necessitating risk adjustment if it is feasible and raising the possibility of either cream skimming or underservice of higher risks if it is not.

In effect, paying providers on some bundled or predetermined basis without adequate risk adjustment adds distorted incentives for those providers to any incentives for insurers to "cream skim." Indeed, even if premiums for competitive insurers were fully risk adjusted, high risks could still suffer provider supply restrictions if they were paid on such a basis *and* their payment was not fully risk adjusted. The possibility of inefficient underservice for high risks and overservice for low risks is a defect of bundled payments (whatever their desirable incentives for cost minimization). Presumably risk adjustment for insurers reduces their incentive to use bundled provider payment (as one of a number of ways of attracting low risks to the insurer) and increases their incentive to use risk adjustment for provider payments if they do use bundled payment (Ma and McGuire, 1997).

7. CONCLUSION

7.1. Risk Variation in a Model of Public Choice: Will the Healthy People Outvote the Sick?

Above, three different types of financing health insurance systems were distinguished:

1. Risk-rated premiums;
2. Community rating with fixed per-capita premiums;
3. Income-related premiums, e.g. a payroll tax.

It was further argued that in a competitive health insurance system, options (2) and (3) require risk adjustment schemes to prevent excessive risk selection. However, it is not clear a priori whether such a system can be politically sustained in a democracy provided that citizen-voters know their risk type and their income-generating capacity. Kifmann (2005) analyzes this question in a model with two-stage voting: in stage 1, the constitutional stage, a public health insurance system of type (2) or (3) is elected if there is unanimous agreement in its favor, and in stage 2, the size of the benefit package is determined by majority rule. The model is based on a number of restrictive assumptions. In particular, premium risk cannot be insured in the private market and at the constitutional stage voters know their income but not their risk

type. There are only two income levels and two risk types and thus four different types of voters.

At the constitutional stage, the poor always prefer the payroll-tax financed system (3) because it redistributes in their favor and eliminates premium risk. For the rich, the second (positive) effect has to be balanced against the first (negative) effect. They may also accept system (3) if their share in total population is larger than the share of high-risk types in the population, if they are sufficiently risk averse, if risk types are sufficiently different and therefore future premiums sufficiently uncertain, and if income inequality as measured by the income ratio is high enough to induce poor and healthy individuals to be in favor of public health insurance but low enough to avoid excessive transfers to the poor.

7.2. Whither Risk Policy?

As we have noted, almost all countries have constructed elaborate mechanisms to control how insurance premiums vary with risks. Those mechanisms in turn affect what insurance plans people choose, and by extension what risk protection they have and what medical expenses they experience. We now know much more about how people respond to such mechanisms, but there is as yet no generally accepted or generally satisfactory method for dealing with the conflicting social goals associated with risk variation.

REFERENCES

Altman, D., Cutler, D., & Zeckhauser, R. (2003). Enrollee mix, treatment intensity, and cost in competing indemnity and HMO plans. *Journal of Health Economics*, *22*, 23–45.

Altman, D., Cutler, D. M., & Zeckhauser, R. J. (1998). Adverse selection and adverse retention. *American Economic Review*, *88*, 122–126.

Armstrong, J., Paolucci, F., McLeod, H., & Van de Ven, W. P. (2010). Risk equalization in voluntary health insurance markets: A three country comparison. *Health Policy*, *98*, 39–49.

Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. *American Economic Review*, *53*, 941–973.

Bajari, P., Hong, H., Khwaja, A., & Marsh, C. (2010). *Moral hazard, adverse selection, and health expenditures: A semiparametric analysis*. Working Paper.

Bauhoff, S. (2010). *Do health plans risk-select? An audit study on Germany's social health insurance*. Working Paper. Harvard University, Cambridge, MA.

Baumann, F., Meier, V., & Werding, M. (2008). Transferable aging provisions in individual health insurance contracts. *German Economic Review*, *9*, 287–311.

Beck, K. & Zweifel, P. (1998). Cream-skimming in deregulated social health insurance: Evidence from Switzerland. In P. Zweifel (Ed.), *Health, the medical profession and regulation* (pp. 211–227). Dordrecht, The Netherlands: Kluwer.

Beck, K., Spycher, S., Holly, A., & Gardiol, L. (2003). Risk adjustment in Switzerland. *Health Policy*, *65*, 63–74.

Bernard, D. & Selden, T. M. (2002). Employer offers, private coverage, and the tax subsidy for health insurance: 1987 and 1996. *International Journal of Health Care Finance and Economics*, *2*, 297–318.

Bhattacharya, J. & Bundorf, M. K. (2009). The incidence of the healthcare costs of obesity. *Journal of Health Economics*, *28*, 649–658.

Bhattacharya, J. & Vogt, W. B. (2006). *Employment and adverse selection in health insurance*. NBER Working Paper 12430. National Bureau of Economic Research, Cambridge, MA.

Biglaiser, G. & Ma, C.-T. A. (2003). Price and quality competition under adverse selection: Market organization and efficiency. *RAND Journal of Economics*, *34*, 266–286.

Blumberg, L. J., Nichols, L. M., & Banthin, J. S. (2001). Worker decisions to purchase health insurance. *International Journal of Health Care Finance and Economics*, *1*, 305–325.

Breyer, F., Heineck, M., & Lorenz, N. (2003). Determinants of health care utilization by German sickness fund members—with application to risk adjustment. *Health Economics*, *12*(5), 367–376.

Brown, J., Duggan, M., Kuziemko, I., & Woolston, W. (2011). *How does risk selection respond to risk adjustment? Evidence from the Medicare Advantage Program*. NBER Working Paper No. 16977. National Bureau of Economic Research, Cambridge, MA.

Brown, R. S., Clement, D. G., Hill, J. W., Retchin, S. M., & Bergeron, J. W. (1993). Do health maintenance organizations work for Medicare? *Health Care Financing Review*, *15*, 7–23.

Browne, M. J. & Doerpinghaus, H. I. (1994–1995). Asymmetric information and the demand for Medigap insurance. *Inquiry*, *31*, 445–450.

Buchmueller, T. & DiNardo, J. (2002). Did community rating induce an adverse selection death spiral? Evidence from New York, Pennsylvania, and Connecticut. *American Economic Review*, *92*, 280–294.

Buchner, F. & Wasem, J. (2003). Needs for further improvement: Risk adjustment in the German health insurance system. *Health Policy*, *65*, 21–35.

Bundorf, M. K. (2002). Employee demand for health insurance and employer health plan choices. *Journal of Health Economics*, *21*, 65–88.

Bundorf, M. K. (2010). The effects of offering health plan choice within employment-based purchasing groups. *Journal of Risk and Insurance*, 77, 105–129.

Bundorf, M. K. & Simon, K. (2006). The impact of rate regulation on access to supplemental health insurance. *American Economic Review Papers and Proceedings*, *96*, 67–71.

Bundorf, M. K., Herring, B. J., & Pauly, M. V. (2010). Health risk, income, and employment-based health insurance. *Forum for Health Economics & Policy*, *13*, Article 13.

Bundorf, M. K., Levin, J. D., & Mahoney, N. (2008). *Pricing and welfare in health plan choice*. NBER Working Paper No. 14153. National Bureau of Economic Research, Cambridge, MA.

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

Cardon, J. H. & Hendel, I. (2001). Asymmetric information in health insurance: Evidence from the national Medicare expenditure survey. *RAND Journal of Economics*, *32*, 408–427.

Carlin, C. & Town, R. (2010). *Adverse selection, welfare and optimal pricing of employer-sponsored health plans*. Working Paper, University of Minnesota, Minneapolis.

Chernew, M. E. & Frick, K. D. (1999). The impact of managed care on the existence of equilibrium in health insurance markets. *Journal of Health Economics*, *18*, 573–592.

Chiappori, P. A. (2000). Econometric models of insurance under asymmetric information. In G. Dionne (Ed.), *Handbook of insurance* (pp. 363–392). Norwell, MA: Kluwer Academic Publishers.

Chiappori, P.-A. & Salanié, B. (2000). Testing for asymmetric information in insurance markets. *Journal of Political Economy*, *108*, 56–78.

Chiappori, P.-A., Jullien, B., Salanié, B., & Salanié, F. (2006). Asymmetric information in insurance: General testable implications. *RAND Journal of Economics*, *37*, 783–798.

Cochrane, J. (1995). Time-consistent health insurance. *Journal of Political Economy*, *103*, 445–473.

Cogan, J. F., Hubbard, R. G., & Kessler, D. P. (2010). The effect of Medicare coverage for the disabled on the market for private insurance. *Journal of Health Economics*, *29*, 418–425.

Cohen, S. B., Ezzati-Rice, T., & Yu, W. (2006). The utility of extended longitudinal profiles in predicting future health care expenditures. *Medical Care*, *44*, I45–I53.

Crocker, K. J. & Snow, A. (1985). The efficiency of competitive equilibria in insurance markets with asymmetric information. *Journal of Public Economics*, *26*, 207–219.

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

Cutler, D. M. & Zeckhauser, R. J. (1998). Adverse selection in health insurance. *Forum for Health Economics & Policy, 1*. (Frontiers in health policy research), Article 2.

Cutler, D. M. & Zeckhauser, R. J. (2000). The anatomy of health insurance. In A. J. Culyer & J. P. Newhouse (Eds.), *Handbook of health economics* (Vol. 1, pp. 563–643). Amsterdam: Elsevier.

Cutler, D. M., Finkelstein, A., & McGarry, K. (2008). Preference heterogeneity and insurance markets: Explaining a puzzle of insurance. *American Economic Review, Papers & Proceedings*, *98*, 157–162.

Davidoff, A., Blumberg, L., & Nichols, L. (2005). State health insurance market reforms and access to insurance for high-risk employees. *Journal of Health Economics*, *24*, 725–750.

Davidson, B. N., Sofaer, S., & Gertler, P. (1992). Consumer information and biased selection in the demand for coverage supplementing Medicare. *Social Science and Medicine*, *34*, 1023–1034.

Del Bene, L. & Vaughan, D. R. (1992). Income, assets, and health insurance: Economic resources for meeting acute health needs of the aged. *Social Security Bulletin*, *55*, 3–25.

Dionne, G. Doherty, N. (1992). Adverse selection in insurance markets: A selective survey. In G. Dionne (Ed.), *Contributions to insurance economics* (pp. 97–140). Boston: Kluwer.

Doiron, D., Jones, G., & Savage, E. (2008). Healthy, wealthy and insured? The role of self-assessed health in the demand for private health insurance. *Health Economics*, *17*, 317–334.

Douven, R. (2007). Morbidity-based risk adjustment in the Netherlands. In E. Wille, V. Ulrich, & U. Schneider (Eds.), *Wettbewerb und Risikostrukturausgleich im internationalen Vergleich* (pp. 161–202). Baden-Baden: Nomos.

Duggan, M., Healy, P., & Morton, F. S. (2008). Providing prescription drug coverage to the elderly: America's experiment with Medicare Part D. *Journal of Economic Perspectives*, *22*, 69–92.

Eckstein, Z., Eichenbaum, M., & Peled, D. (1985). Uncertain lifetimes and the welfare enhancing properties of annuity markets and social security. *Journal of Public Economics*, *26*, 303–320.

Eggleston, K. (2000). Risk selection and optimal health insurance-provider payment systems. *Journal of Risk and Insurance*, *67*, 173–196.

Ehrlich, I. & Becker, G. S. (1972). Market insurance, self-insurance, and self-protection. *Journal of Political Economy, 80*, 623–648.

Eichner, M. J., McClellan, M. B., & Wise, D. (1997). Health expenditure, persistence, and the feasibility of medical savings accounts. *Tax Policy and the Economy, 11*, 91–128.

Einav, L., Finkelstein, A., & Cullen, M. R. (2010a). Estimating welfare in insurance markets using variation in prices. *Quarterly Journal of Economics*, *125*, 877–921.

Einav, L., Finkelstein, A., & Levin, J. (2010b). Beyond testing: Empirical models of insurance markets. *Annual Review of Economics*, *2*, 311–336.

Einav, L., Finkelstein, A., & Schrimpf, P. (2007). *The welfare cost of asymmetric information: Evidence from the U.K. annuity market*. NBER Working Paper 13228. National Bureau of Economic Research, Cambridge, MA.

Ellis, R. P. & McGuire, T. G. (2007). Predictability and predictiveness in health care spending. *Journal of Health Economics*, *26*, 25–48.

Enthoven, A. (1988). *Theory and practice of managed competition in health care finance*. Amsterdam: North-Holland.

Enthoven, A. & Kronick, R. (1989). A consumer-choice health plan for the 1990s. *New England Journal of Medicine*, *320*, 29–37.

Ettner, S. L. (1997). Adverse selection and the purchase of Medigap insurance by the elderly. *Journal of Health Economics*, *116*, 543–562.

Fama, T., Fox, P. D., & White, L. A. (1995). HMOs care for the chronically ill? *Health Affairs*, *14*, 234–243.

Fang, H., Keane, M. P., & Silverman, D. (2008). Sources of advantageous selection: Evidence from the Medigap insurance market. *Journal of Political Economy, 116*, 303–350.

Finkelstein, A. & McGarry, K. (2006). Multiple dimensions of private information: Evidence from the long-term care insurance market. *American Economic Review, 96*, 938–958.

Finkelstein, A., Poterba, J., & Rothschild, C. (2006). *Redistribution by insurance market regulation: Analyzing a ban on gender-based retirement annuities*. NBER Working Paper 12205. National Bureau of Economic Research, Cambridge, MA.

Florence, C. S. & Thorpe, K. E. (2003). How does the employer contribution for the federal employees' health benefits program influence plan selection? *Health Affairs*, *22*, 211–218.
Frank, R. G., Glazer, J., & McGuire, T. G. (2000). Measuring adverse selection in managed health care. *Journal of Health Economics*, *19*, 829–854.
Garber, A. M., MaCurdy, T. E., & McClellan, M. B. (1998). Persistence of medical expenditures among elderly beneficiaries. *Frontiers in Health Policy Research*, *1*, 153–180.
Glazer, J. & McGuire, T. G. (2000). Optimal risk adjustment in markets with adverse selection: An application to managed care. *American Economic Review*, *90*, 1055–1071.
Glazer, J. & McGuire, T. G. (2002). Setting health plan premiums to ensure efficient quality in health care: Minimum variance, optimal risk adjustment. *Journal of Public Economics*, *84*, 153–175.
Glazer, J. & McGuire, T. G. (2009). *Gold and silver health plans: Accommodating demand heterogeneity in managed competition*. Working Paper, Boston, MA.
Glied, S. A. (2005). *The employer-based health insurance system: Mistake or cornerstone? Policy challenges in modern health care*. Rutgers University Press. Available at: <http://www.rwjf.org/pr/product.jsp?id=14805/>.
Goldman, D. (1995). Managed care as a public cost containment mechanism. *RAND Journal of Economics*, *26*, 277–295.
Goldstein, G. S. & Pauly, M. V. (1976). Group health insurance as a local public good. In R. N. Rosett (Ed.), *The role of health insurance in the health services sector* (pp. 73–114). New York: National Bureau of Economic Research.
Gruber, J. (1994). The incidence of mandated maternity benefits. *American Economic Review*, *84*, 622–641.
Gruber, J. (2000). Health insurance and the labor market. In A. J. Culyer & J. P. Newhouse (Eds.), *Handbook of health economics* (Vol. 1, pp. 645–706). Amsterdam: Elsevier.
Handel, B. R. (2010). Essays in consumer behavior and market outcomes in health insurance. Dissertation, Department of Economics, University of California at Berkeley.
Harsanyi, J. (1955). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy*, *63*, 309–321.
Helderman, J., Schut, F., Van der Grinten, T., & Van de Ven, W. (2005). Market-oriented health care reforms and policy learning in the Netherlands. *Journal of Health Politics, Policy and Law*, *30*, 189–209.
Hellinger, F. J. (1987). Selection bias in health maintenance organizations: Analysis of recent evidence. *Health Care Financial Review*, *9*, 55–63.
Hellinger, F. J. (1995). Selection bias in HMOs and PPOs: A review of the evidence. *Inquiry*, *32*, 135–142.
Hellinger, F. J. & Wong, H. S. (2000). Selection bias in HMOs: A review of the evidence. *Medical Care Research and Review*, *57*, 405–439.
Hellwig, M. (1987). Some recent developments in the theory of competition in markets with adverse selection. *European Economic Review*, *31*, 319–325.
Herring, B. & Pauly, M. V. (2006). Incentive-compatible guaranteed renewable health insurance. *Journal of Health Economics*, *25*, 395–417.
Herring, B. J. & Pauly, M. V. (2007). The demand for health insurance in the group setting: Can you always get what you want? *Journal of Risk and Insurance*, *74*, 115–140.
Hirshleifer, J. (1971). The private and social value of information and the reward to inventive activity. *American Economic Review*, *61*, 561–574.
Hsu, J., Fung, V., Huang, J., Price, M., Brand, R., Hui, R., et al. (2010). Fixing flaws in Medicare drug coverage that prompt insurers to avoid low-income patients. *Health Affairs*, *29*, 2335–2343.
Hsu, J., Huang, J., Fung, V., Price, M., Brand, R., Hui, R., et al. (2009). Distributing $800 billion: An early assessment of Medicare Part D risk adjustment. *Health Affairs*, *28*, 215–225.
Jack, W. (2006). Optimal risk adjustment with adverse selection and spatial competition. *Journal of Health Economics*, *25*, 908–926.
Keenan, P. S., Buntin, M. J., McGuire, T. G., & Newhouse, J. P. (2001). The prevalence of formal risk adjustment in health plan purchasing. *Inquiry*, *38*, 245–259.
Kifmann, M. (2005). Health insurance in a democracy: Why is it public and why are premiums income related? *Public Choice*, *124*, 283–308.

Kifmann, M. (2006). Risk selection and complementary health insurance: The Swiss approach. *International Journal of Health Care Finance and Economics*, *6*, 151–170.
Kifmann, M. & Roeder, K. (2010). *Can risk-based premiums be more equitable than social insurance?* Working Paper, University of Augsburg.
Kronick, R. & Dreyfus, T. (1996). Diagnostic risk adjustment for Medicaid: The disability payment system. *Health Care Financing Review*, *17*, 7–33.
Lamers, L. M. (2001). Health-based risk adjustment: Is inpatient and outpatient diagnostic information sufficient? *Inquiry*, *38*, 423–431.
Lamers, L. M., Van Vliet, R., & Van de Ven, W. (2003). Risk adjusted premium subsidies and risk sharing: Key elements of the competitive sickness fund market in the Netherlands. *Health Policy*, *65*, 49–62.
Lo Sasso, A. T. & Lurie, I. Z. (2009). Community rating and the market for private non-group health insurance. *Journal of Public Economics*, *93*, 264–279.
Luft, H. S. & Dudley, R. A. (2003). Measuring quality in modern managed care. *Health Services Research*, *38*, 1373–1379.
Ma, C.-T. A. & McGuire, T. M. (1997). Optimal health insurance and provider payment. *American Economic Review*, *87*, 685–704.
McGuire, T. G., Newhouse, J. P., & Sinaiko, A. D. (2011). An economic history of Medicare Part C. *Milbank Quarterly*, *89*, 289–332.
Mello, M. M., Stearns, S. C., Norton, E. C., & Ricketts, T. C., III (2003). Understanding biased selection in Medicare HMOs. *Health Services Research*, *38*, 961–992.
Miller, N. H. (2005). Pricing health benefits: A cost minimization approach. *Journal of Health Economics*, *24*, 931–949.
Miyazaki, H. (1977). The rat race and internal labor markets. *Bell Journal of Economics*, *8*, 394–418.
Monheit, A. C. (2003). Persistence in health expenditures in the short run: Prevalence and consequences. *Medical Care*, *41*, III53–III64.
Monheit, A. C. & Steinberg Schone, B. (2004). How has small group reform affected employee health insurance coverage? *Journal of Public Economics*, *88*, 237–254.
Monheit, A. C. & Vistnes, J. P. (1994). Implicit pooling of workers from large and small firms. *Health Affairs*, *13*, 301–314.
Moran, J. R., Chernew, M. E., & Hirth, R. A. (2001). Preference diversity and the breadth of employee health insurance options. *Health Services Research*, *36*, 911–934.
Newhouse, J. P. (1996). Reimbursing health plans and health providers: Efficiency in production versus selection. *Journal of Economic Literature*, *34*, 1236–1263.
Newhouse, J. P., Manning, W. G., Keeler, E. B., & Sloss, E. M. (1989). Adjusting capitation rates using objective health measures and prior utilization. *Health Care Financing Review*, *10*, 41–54.
Nicholson, S., Bundorf, M. K., Stein, R. M., & Polsky, D. (2004). The magnitude and nature of risk selection in employer-sponsored health plans. *Health Services Research*, *39*, 1817–1838.
Pauly, M. & Herring, B. (1999). *Pooling health insurance risks*. Washington, DC: AEI Press.
Pauly, M. V. (1968). The economics of moral hazard. *American Economic Review*, *58*, 531–537.
Pauly, M. V. (1974). Overinsurance and public provision of insurance: The roles of moral hazard and adverse selection. *Quarterly Journal of Economics*, *88*, 44–62.
Pauly, M. V. & Herring, B. (2001). Expanding insurance coverage through tax credits: Tradeoffs and options. *Health Affairs*, *20*, 1–18.
Pauly, M. V. & Herring, B. (2007). Risk pooling and regulation: Policy and reality in today's individual health insurance market. *Health Affairs*, 26(3), 770–779.
Pauly, M. V. & Lieberthal, R. (2008). How risky is individual health insurance? *Health Affairs*, *27*, w242–w249 (published online May 6, 2008; 10.1377/hlthaff.27.3.w242).
Pauly, M. V. & Zeng, Y. (2004). Adverse selection and the challenges to stand-alone prescription drug insurance. In D. M. Cutler & A. M. Garber (Eds.), *Frontiers in health policy research* (Vol. 7, pp. 55–74). Cambridge, MA: National Bureau of Economic Research and the Massachusetts Institute of Technology.
Pauly, M. V., Danzon, P., Feldstein, P., & Hoff, J. (1992). *Responsible national health insurance*. Washington, DC: AEI Press.

Pauly, M. V., Kunreuther, H., & Hirth, R. (1995). Guaranteed renewability in insurance. *Journal of Risk and Uncertainty, 10*, 143–156.
Polsky, D. & Nicholson, S. (2004). Why are managed care plans less expensive? Risk selection, utilization, or reimbursement? *Journal of Risk and Insurance, 71*, 21–40.
Pope, G. C., Kautter, J., Ellis, R. P., Ash, A. S., Ayanian, J. Z., Iezzoni, L. I., et al. (2004). Risk adjustment of Medicare capitation payments using the CMS-HCC model. *Health Care Financing Review, 25*, 119–141.
Prinsze, F. J. & Van Vliet, R. C. (2007). Health-based risk adjustment: Improving the pharmacy-based cost group model by adding diagnostic cost groups. *Inquiry, 44*, 469–480.
Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Harvard University Press.
Robinson, J. C. & Gardner, L. B. (1995). Adverse selection among multiple competing health maintenance organizations. *Medical Care, 33*, 1161–1175.
Robinson, J. C., Gardner, L. B., & Luft, H. S. (1993). Health plan switching in anticipation of increased medical care utilization. *Medical Care, 31*, 43–51.
Robst, J. (2006). Estimation of a hedonic pricing model for Medigap insurance. *Health Services Research, 41*, 2097–2113.
Robst, J., Levy, J. M., & Ingber, M. J. (2007). Diagnosis-based risk adjustment for Medicare prescription drug plan payments. *Health Care Financing Review, 28*, 15–30.
Rothschild, M. & Stiglitz, J. (1976). Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *Quarterly Journal of Economics, 90*, 629–650.
Schaefer, E. & Reschovsky, J. D. (2002). Are HMO enrollees healthier than others? Results from the community tracking study. *Health Affairs, 21*, 249–258.
Selden, T. M. (1999). Premium subsidies for health insurance: Excessive coverage vs. adverse selection. *Journal of Health Economics, 18*, 709–725.
Shen, Y. & Ellis, R. P. (2002). Cost-minimizing risk adjustment. *Journal of Health Economics, 21*, 515–530.
Shewry, S., Hunt, S., Ramey, J., & Bertko, J. (1996). Risk adjustment: The missing piece of market competition. *Health Affairs, 15*, 171–181.
Shmueli, A. (2001). The effect of health on acute care supplemental insurance ownership: An empirical analysis. *Health Economics, 10*, 341–350.
Short, P. F. & Vistnes, J. P. (1992). Multiple sources of Medicare supplementary insurance. *Inquiry, 29*, 33–43.
Simon, K. I. (2005). Adverse selection in health insurance markets? Evidence from state small-group health insurance reforms. *Journal of Public Economics, 89*, 1865–1877.
Song, Y., Skinner, J., Bynum, J., Sutherland, J., Wennberg, J. E., & Fisher, E. S. (2010). Regional variations in diagnostic practices. *New England Journal of Medicine, 363*, 45–53.
Spence, A. M. (1978). Product differentiation and performance in insurance markets. *Journal of Public Economics, 10*, 427–447.
Strohmenger, R. & Wambach, A. (2000). Adverse selection and categorical discrimination in the health insurance markets: The effects of genetic tests. *Journal of Health Economics, 19*, 197–218.
Subramanian, K., Lemaire, J., Hershey, J. C., Pauly, M. V., Armstrong, K., & Asch, D. A. (1999). Estimating adverse selection costs from genetic testing for breast and ovarian cancer: The case of life insurance. *Journal of Risk and Insurance, 66*, 531–550.
Tabarrok, A. (1994). Genetic testing: An economic and contractarian analysis. *Journal of Health Economics, 13*, 75–91.
US General Accounting Office (GAO) (2002). *MEDIGAP: Current policies contain coverage gaps, undermine cost control incentives*. GAO-02-533T, Washington, DC.
Van de Ven, W. (2006). Response: The case for risk-based subsidies in public health insurance. *Health Economics, Policy and Law, 1*, 195–199.
Van de Ven, W. & Ellis, R. (2000). Risk adjustment in competitive health plan markets. In A. J. Culyer & J. P. Newhouse (Eds.), *Handbook of health economics* (Vol. 1, pp. 755–845). Amsterdam: Elsevier.
Van de Ven, W. & Schut, F. (2008). Universal mandatory health insurance in the Netherlands: A model for the United States? *Health Affairs, 27*, 771–781.

Van de Ven, W. & Van Vliet, R. (1992). How can we prevent cream skimming in a competitive health insurance market? In P. Zweifel & H. Frech, III (Eds.), *Health economics worldwide* (pp. 23–46). Dordrecht: Kluwer.

Van de Ven, W., Beck, K., Buchner, F., Chernichovsky, D., Gardiol, L., Holly, A., et al. (2003). Risk adjustment and risk selection on the sickness fund insurance market in five European countries. *Health Policy, 65*, 75–98.

Van Vliet, R. (1992). Predictability of individual health care expenditures. *Journal of Risk and Insurance, 59*, 443–460.

Van Vliet, R. (2006). Free choice of health plan combined with risk-adjusted capitation payments: Are switchers and new enrolees good risks? *Health Economics, 15*, 763–774.

Vera-Hernández, M. (2003). Structural estimation of a principal–agent model: Moral hazard in medical insurance. *RAND Journal of Economics, 34*, 670–693.

Vistnes, J. P. & Banthin, J. S. (1997–1998). The demand for Medicare supplemental insurance benefits: The role of attitudes toward medical care and risk. *Inquiry, 34*, 311–324.

Wagstaff, A. & Van Doorslaer, E. (2000). Equity in health care finance and delivery. In A. J. Culyer & J. P. Newhouse (Eds.), *Handbook of health economics* (Vol. 1, pp. 1803–1862). Amsterdam: Elsevier.

Welch, W. P. (1985). Medicare capitation payments to HMOs in light of regression toward the mean in health care costs. *Advances in Health Economics and Health Services Research, 6*, 75–100.

Wilson, C. (1977). A model of insurance markets with incomplete information. *Journal of Economic Theory, 16*, 167–207.

Wolfe, J. R. & Goddeeris, J. H. (1991). Adverse selection, moral hazard, and wealth effects in the Medigap insurance market. *Journal of Health Economics, 10*, 433–459.

Zuvekas, S. H. & Cohen, J. W. (2007). Prescription drugs and the changing concentration of health care expenditures. *Health Affairs, 26*, 249–257.

Zweifel, P. & Breuer, M. (2006). The case for risk-based premiums in public health insurance. *Health Economics, Policy and Law, 1*, 171–188.