CHAPTER 17, Handbook of Health Economics (eds. A.J. Culyer and J.P. Newhouse)

RISK ADJUSTMENT IN COMPETITIVE HEALTH PLAN MARKETS

Wynand P.M.M. van de Ven
Erasmus University Rotterdam

Randall P. Ellis
Boston University

March 31, 1999

JEL-classification codes: C10, D82, G22, I10, I11, I18

Acknowledgements
The authors gratefully acknowledge very useful comments from: Arlene Ash, Ernie Berndt, Richard Frank, Jeremiah Hurley, Leida Lamers, Tom McGuire, Joe Newhouse, Erik Schokkaert, Matthias von der Schulenburg, Erik van Barneveld, René van Vliet and Peter Zweifel.

Contents

1. Introduction

More so than any other good routinely used by consumers, health care expenditures are characterized both by large *random* variation as well as large *predictable* variation across individuals. Such differences create the potential for large efficiency gains due to risk reduction from insurance, and raise important concerns about fairness across individuals with different expected needs for services. In this chapter, we examine the principles and practice of risk adjustment and how it may contribute to both efficiency and equity in competitive health plan markets.

Because the term "risk adjustment" is used in different contexts to mean different things, we begin by defining how we shall use the term. Throughout this chapter we use *risk adjustment* to mean the use of information to calculate the expected health expenditures of individual consumers over a fixed interval of time (e.g., a month, quarter, or year) and set subsidies to consumers or health plans to improve efficiency and equity. By this definition we intend to exclude the use of risk modeling for profiling, or measuring resources defined over episodes of treatment or episodes of illness [See Iezzoni (1994) for discussion of this practice], which is also known as severity adjustment. We also exclude the adjustment of expected expenditures at the family, group, or plan level, such as is commonly done by actuaries using occupational and demographic averages. Although risk adjusters may be used by insurers for risk-rating their premiums, we do not focus on this application. Risk adjusters may also be used for monitoring, or for internal financing decisions within managed care organizations (e.g., risk-adjusted capitation payments and shared risk pools), or included as control variables in prediction models with other objectives, but these uses are not the focus of our chapter.

As our title indicates, we focus our discussion on risk adjustment in the context of competitive health plan markets. By *competitive,* we mean markets in which individual consumers have a periodic choice of health plan and health plans may take actions, such as designing, pricing and marketing their products, to attract or repel enrollees. By *health plan* we mean a risk-bearing entity that performs at least some insurance function - i.e. it bears some or all of the financial risk associated with the random variation in health expenditures across individuals. Health plans may also manage or provide health care, and this can influence how risk-adjusted payments should be made, however we focus primarily on plan-level rather than provider-level incentives. Examples of health plans are: private health insurance companies, sickness funds (Israel, Netherlands), managed care organizations like Health Maintenance Organizations (US) and capitated provider groups like general practitioner-fundholders (U.K.).

1.1. Efficiency and fairness

Imperfect information is a serious problem in health plan markets. Yet efficiency and equity issues would need to be addressed even in a world with *perfect information,* since plans will face large differences in expected health costs due to heterogeneity in demographics and the incidence of illness. A competitive market forces health plans to break even, in expectation, on each insurance contract offered. If a health plan does not adjust its premium for a risk factor that is known either to individuals or to plans, then low-risk individuals will tend to choose a competing plan that offers a lower premium or a contract specifically designed to attract low risk individuals. Consequently the first plan, left with only high-risk individuals, will have to increase its premium. In this way, in the absence of any restrictions on premium rates a competitive health plan market will tend to result in plans charging risk-adjusted premiums that differentiate according to the individual consumer's risk. This is called the equivalence principle.[1] Risk-adjusted premiums are the norm, not the exception, in competitive markets, and in the absence of regulation, health plans will tend to charge premiums that differ across both observable risk factors and benefit packages designed to attract specific risk types.

This raises the equity question: is this fair? As we document below, risk-adjusted premiums can easily differ by a factor of ten or more for demographic risk factors such as age, and factors of 100 or more once health status is also taken into account. Almost universally, people agree that premiums which reflect such large differences are not fair, and that cross subsidies are needed.

[1] We assume that, up to a sophisticated level of risk-rating, the costs of risk-rating are not prohibitively high. If risk-rating becomes too costly, technically infeasible, or politically unacceptable, the equivalence principle may force health plans to exclude from coverage the costs related to some preexisting medical conditions or to refuse to contract with high-risk individuals altogether.

In addition to the equity concerns there are also efficiency problems: consumers are not permitted to equalize the marginal utility of income across different annual or lifetime health profiles. Risk averse consumers would like to buy insurance against the risk of becoming a bad risk in the future. However, in practice there is no market for such insurance. The welfare losses resulting from this inefficiency[2] of a competitive health plan market are discussed in chapter XXX (this handbook) by Cutler and Zeckhauser [see also Newhouse (1984); Pauly (1992); Diamond (1992); Cochrane (1995)].

Problems are exacerbated if there is *asymmetric information*, with consumers knowing more than health plans. This asymmetry can create moral hazard and adverse selection inefficiencies. Consider the moral hazard problem that arises when consumers have private information about their health care needs which is not known to the health plan. If consumers are fully insured against financial risks, then they will tend to over consume health services because of the moral hazard problem [Arrow (1963); Chapter XXX (this handbook) by Zweifel and Manning, and Chapter XXX (this handbook) by Cutler and Zeckhauser]. To reduce this problem, health plans typically try to constrain the use of services through demand-side incentives (such as user fees, deductibles, copayments, waiting time, etc.) or supply incentives (supply-side cost sharing, case management, selection of providers, etc.). Unfortunately, the same tools that health plans use to offset the patient-level moral hazard problem can also be used to compete for profitable enrollees: competing health plans will design their plans so as to attract a favorable selection of enrollees.[3]

1.2. The supply price and demand price of insurance

The payment received by a health plan for an individual enrollee need not be the same as the payment made by that same enrollee: the supply price and the demand price for health insurance can differ. This important distinction is often missed. Note that we refer here to the health plan price, commonly called the insurance premium, not the price paid at the time health services are received. Subscribers rarely pay the full insurance premium. Instead, with only a few exceptions, a substantial part of the insurance premiums tend to be paid by a *sponsor*.[4] The sponsor acts as a broker in structuring coverage, contracting with and regulating health plans, and managing enrollment. The sponsor also reallocates the burden of health plan premiums across consumers, and enters into risk-sharing arrangements with health plans [cf. Enthoven (1988)]. The demand price and the supply price will differ only if a sponsor redistributes the financial burden.

The sponsor can be of many types - an employer, a coalition of employers, a government agency, a nonprofit organization, or a distinct insurance entity empowered to use coercion to redistribute risk. Examples of sponsors are the Health Care Financing Administration in the US which negotiates "at-risk" contracts with HMOs for Medicare beneficiaries and the government agencies which regulate and even pay the competitive sickness funds in several European countries. In many countries, the sponsor role is fulfilled by the government agency that regulates access to individual (or small group) private health insurance coverage in a competitive market.[5] In the US, the role of sponsor is also fulfilled by (large) employers who offer group health insurance to their employees.[6]

There is no widely used terminology for distinguishing the demand and supply prices for health plans, so we define our own. On the demand side, we call payments made by the consumer *contributions,* the two most important of which are *premium contributions* - the contribution of a consumer towards his own health insurance coverage - and *solidarity contributions,* which are made toward all consumers covered by the sponsor (see Figure 1). The term solidarity contribution derives from a substantial literature in Europe on the 'solidarity principle', which holds that high-risk individuals should receive a subsidy to increase their access to health insurance coverage [see e.g.,

[2] The fairness issue discussed above can alternatively be thought of as an inefficiency because there is no market for buying insurance against a bad draw from the gene pool (i.e. lifetime insurance).

[3] A point we develop more fully below is that this selection can arise either because of asymmetric information, or because of regulation-induced pooling of people with different known risks.

[4] Newhouse (1996) calls the sponsor the regulator. We use sponsor to highlight the redistribution role, not just the fact that the sponsor may also regulate the characteristics of health plans that are offered.

[5] For example, in Australia, Chile, Ireland, the Netherlands, Portugal, Spain and the US.

[6] In a later section of this chapter, we discuss the implications of the sponsor being voluntarily chosen by consumers, but for the most part we focus on the common case in which there is no consumer choice of sponsor.

Hamilton (1997), and Chinitz et al. (1998)].[7] On the supply side, we call the payments made by the sponsor *subsidies*.[8] The most important type of sponsor subsidy is the *premium subsidy*, an ex-ante subsidy mostly paid directly to the health plan.[9] The sum total of ex-ante payments received by the health plan for one consumer, i.e. the premium contribution plus the premium subsidy, is the (supply side) *health plan premium*, or simply the *premium*. As discussed below, a wide variety of mechanisms are used for calculating the consumer contributions and the premium subsidies, as well as for organizing the actual payment flows in practice.

Figure 1

1.3. The role of the sponsor

The sponsor plays a crucial role in enabling health plan premiums to be risk-adjusted (reflecting the expected health cost of the plans' enrollees) while not insisting that payments by individuals reflect each person's own expected cost. One mechanism for doing this is to *risk adjust* the premium subsidies to competing health plans while charging consumers a solidarity contribution that does not reflect the person's own expected cost. Another mechanism to reduce the variation in contributions across consumers is to *regulate* the rate classes, plan features, and premium contributions that health plans are allowed to charge. As we highlight below, it is difficult for the sponsor to fully risk adjust health plans subsidies, but it is also difficult to fully regulate all of the dimensions in which health plans will try to differentiate their plan features.

Sponsors have many mechanisms for allocating financial burdens among consumers through the contribution side of the market, as well as great flexibility in redistributing financial revenues among health plans on the supply side of the market. Note that once the linkage between the contribution and expected health care use is broken by the sponsor, then solidarity contributions can be based on information that may have little relation to future health costs, such as income. It is common for solidarity contributions to be income-based, or to be a flat payment that does not vary across plans with different benefit designs. Throughout this chapter, we focus primarily on so-called "risk-solidarity", that is solidarity between high- and low-risk individuals. Solidarity between high- and low-income individuals, so-called "income-solidarity", is a redistribution concept that varies across countries and is relatively independent of the incentive issues and fairness across risk types that is the primary focus here.

1.4. Policy relevance

The policy relevance of an adequate risk adjustment mechanism has increased during the 1990s as many countries make their individual health insurance market more competitive or reform their already competitive markets in order to increase access to coverage for high-risk individuals.[10] Many countries have chosen to use prospective payment arrangements (pure or otherwise) for health plans as a means for creating incentives to be cost conscious, together with competition among health plans as a tool for preserving quality, innovation and responsiveness to consumer preferences. Risk adjustment is a key strategy for attenuating problems that threaten the effectiveness of this strategy for resource allocation in health care. Without adequate risk adjustment it is hard, if not impossible, to achieve both efficiency and fairness objectives in a competitive health plan market.[11].

Despite its increasing relevance, the practical application of risk adjustment is still at early stages. For reasons that are not clear to us, most sponsors around the world do not use risk adjustment. Instead, they regulate the dimensions along which health plans are allowed to compete. They force plans to pool consumers into a relatively small number

[7] In the US there does not appear to be any widely-used terminology for describing the normative concept that high-risk individuals should receive a cross-subsidy from low-risk individuals.
[8] Researchers in the US are more used to thinking of the employer as being the sponsor, and focusing on employee and employer contributions toward the premium. This terminology ignores the fact that the sponsor need not to be an employer, and that the consumers may make other payments besides the premium contributions.
[9] Another type of sponsor subsidy is the ex-post payments made by the sponsor to the health plans because of the risk-sharing arrangements between the sponsor and the health plans (see section 4).
[10] For example Belgium, Colombia, the Czech Republic, Germany, Ireland, Israel, The Netherlands, Poland, Russia, Switzerland and the US.
[11] Risk adjustment is also relevant for a competitive provider market where risk-adjusted payments are used - often by a large monopsonistic insurer (e.g., a governmental agency) - to push financial risks all the way down to providers. For example, in the GP fundholder system of the UK, primary care physicians receive a risk-adjusted capitation payment for some or all of the follow-up care of their patients. In the terminology of this chapter, such a GP fund holder is considered a health plan.

of rate categories and regulate the characteristics of contracts offered to each of these categories.[12]

Whereas a system of risk-adjusted subsidies attempts to provide *explicit* subsidies to high-risk individuals, the effect of regulating plan design and restricting the variation of premium contributions is to create *implicit* cross-subsidies from low-risk to high-risk individuals. Although this risk pooling may foster the solidarity principle, it creates *predictable* losses for health plans on their high-risk individuals. In so doing, it creates incentives for health plans to avoid individuals with predictable losses and to select predictably profitable insureds.[13] This selection and the resulting risk segmentation can adversely affect access to care, quality of care and efficiency (see section 2.5.).

If premium subsidies cannot be adequately risk adjusted or if loosening the restrictions on the variation of the premium contributions is not socially acceptable, the adverse effects of selection may also be reduced by various forms of *ex post* risk sharing between the sponsor and the health plans. *Risk sharing* implies that the health plans are retrospectively reimbursed by the sponsor for some of their costs. Although risk sharing effectively reduces the health plans' incentives for selection, it also reduces their incentives for efficiency [Newhouse (1996)].

The conclusion is that in competitive health plan markets - given that risk-adjusted subsidies will always be imperfect – there will always be selection incentives. Because the effects of selection have consequences for both efficiency and fairness, we are confronted with a complicated *tradeoff between efficiency and fairness objectives.* The relevance of an adequate risk adjustment mechanism is that the better the explicit subsidies are adjusted for relevant risk factors, the less severe is the tradeoff. In theory, perfect risk adjustment can eliminate this tradeoff entirely.

1.5. Outline

This chapter gives an overview of all aspects of risk adjustment in competitive health plan markets. We also discuss at length the major mechanisms that can be either a complement or an alternative to risk adjustment, namely plan regulation, carveouts, and ex post risk sharing. The chapter is relevant for voluntary health plan markets as well as for mandatory health plan membership.

The organization of this chapter is as follows. Section 2 presents a conceptual framework of risk adjustment. Section 3 extensively discusses the state of the art of empirical *risk adjusters*, i.e. the predictors used in risk adjustment. Section 4 discusses several forms of risk sharing, which can be used as a tool for reducing selection. The practice of risk adjustment and risk sharing in several countries is discussed in section 5. Finally some directions for future research are discussed in section 6.

2. Conceptual aspects of risk adjustment

In subsection 2.1 we briefly consider each of the three payment flows identified in the preceding section: risk-adjusted premium subsidies, solidarity contributions, and premium contributions. In section 2.2 we discuss some conceptual aspects of how to calculate the risk-adjusted subsidies. In subsections 2.3-2.6 we discuss the consequences of regulations that sponsors may implement as a substitute for, or as a complement to, risk adjustment.[14]

2.1. Payment flows

2.1.1. Risk-adjusted premium subsidies

The central feature of any risk adjustment system is a risk-adjusted premium subsidy (or voucher) from the sponsor to each consumer or to high-risk consumers only. In most countries the sponsor pays the subsidy directly to the

[12] Although these regulations reduce the ability of plans to select profitable enrollees, they increase the incentive for health plans to try to do so.
[13] If the sponsor (e.g., an employer) contracts with only one health plan, risk adjustment is not needed to prevent selection by the plan. However, if the single health plan offers its beneficiaries a menu of several options to choose among, selection may occur *within* the health plan. Even if there is only one plan and no choice by enrollees, risk adjustment may still be used within the health plan to allocate payments among providers. We do not focus attention on how risk adjustment may be used for these internal financing decisions within health plans.

[14] Section 2 is partly based on Van de Ven et al. (1997).

consumer's health plan and thereby lowers the consumer's premium contribution (see Figure 1). The risk-adjusted premium subsidy has several general properties that are worth highlighting. The subsidy is generally worth a specified amount of money, dependent only on the individual's relevant risk characteristics.[15] We assume that the subsidy does not depend on the premium that the consumer pays or the specific health plan chosen by the consumer. The subsidy may be earmarked for the purchase of a specified health plan with specified coverage features, or may be portable across plans.[16] The risk-adjusted subsidy is not transferable. The information that may be used by the sponsor to calculate the risk-adjusted subsidy is discussed in section 2.2 and section 3.

2.1.2. Solidarity contributions

Solidarity contributions are payments made by consumers toward the health needs of everyone covered by the sponsor, not payments made for a consumer's own health care. Such payments may reflect information that is largely unrelated to the individual's health care needs (income, or wealth). Solidarity contributions are mandatory payments by enrollees, made independently of the plan or benefit features selected.

Although in Figure 1 for simplicity we show the premium subsidies as financed entirely by mandatory solidarity contributions from enrollees, the sponsor's outlay may also include financing from other sources. In the US the risk-adjusted subsidies to HMOs with Medicare risk contracts (based on an Adjusted Average Per Capita Cost calculation) are financed primarily out of federal payroll taxes. In the Netherlands the risk-adjusted subsidies are supported from a combination of earmarked income-related enrollees' contributions, general taxes, and a mandatory levy on the premium of each private health insurance contract.

In some countries, such as the US, some individuals get to choose their sponsor when they change employment (e.g., their employer, or a sponsor for the unemployed). When solidarity contributions or premium subsidies differ across sponsors for identical plans, then individuals have an incentive to select a sponsor that contributes more generously. Such differences can also make enrollees reluctant to leave a sponsor with a favorable solidarity contribution. For example, in the US, unemployed persons often have more generous coverage through the Medicaid programs than do low-wage workers. See Chapter XXX in this volume by Jon Gruber for a discussion of distortion in labor markets resulting from concerns about loss of sponsorship.

2.1.3. Premium contributions

A premium contribution by an enrollee is a payment for his or her own health plan. A consumer's premium contribution equals the health plan's premium minus the premium subsidy. Differences in expected costs across individuals may be reflected either in differences in premium subsidies or in differences in premium contributions. If the premium subsidies are adjusted for differences in health status across individuals, the premium contribution will be unrelated to an enrollee's health status. If the premium subsidies are not adjusted for differences in plan benefit features or efficiency of provision, these differences in expected costs will typically be reflected in the premium contributions.

In the sickness fund system in the Netherlands the risk-adjusted subsidy equals the risk-adjusted predicted per capita costs at the national level minus a fixed amount that is identical for all persons. In the US Medicare system, risk-contracting HMOs are paid 95 percent of the risk-adjusted predicted per capita costs. In both countries the health plans are allowed to make up for any potential shortfall in this premium subsidy by charging a *community-rated* (i.e. the same) premium contribution to all enrollees who choose the same plan.[17] Each health plan is free to set its own premium contribution. In the Netherlands in 1999 the premium contributions varied between 345 and 441 Dutch guilders per enrollee per year.[18] In the US in 1996 63 per cent of Medicare risk-contract enrollees were quoted a zero premium contribution. The other 37 per cent of enrollees paid an average premium contribution of 162 US $ per enrollee per year

[15] That is the risk factors for which solidarity is desired (see section 2.2).
[16] The sponsor may define a *minimum* benefits package which health plans may extend with additional benefits (like e.g., the US Medicare risk contracts) or the sponsor may require all health plans to offer a fully *standardized* benefits package (as in the Dutch sickness fund system). The advantage of a minimum package is that health plans can be responsive to consumer preferences (no one-size-fits-all coverage). Disadvantages of a minimum package are that (1) the benefits package can be used as a tool for cream skimming; (2) it reduces the transparency of health plan products; (3) it reduces the price competition because of segmentation of the market.
[17] In the specific case of community-rated premium contributions the premium subsidy is often referred to as *"capitation"*.
[18] i.e. between 172 and 220 US$ per year (1999 exchange rate).

[Lamphere et al. (1997)]. In other countries, (e.g., Israel and Colombia) the sponsor requires the premium contribution to be zero for all enrollees. That is, the health plan premiums equal the risk-adjusted premium subsidies.

2.1.4. Different modalities of payment flows

Figure 1 shows schematically how the risk adjustment system is applied in Medicare in the US and the sickness fund system in the Netherlands. We refer to such an implementation as modality A. However, actual payment flows in a risk adjustment system need not follow this pattern. One alternative is that the premium subsidies go to the consumer, who then pays the total premium directly to the health plan, (a so-called "voucher model"). A second alternative is that the sponsor also collects the premium contributions and transfers them to the health plans. This alternative is applied by some employer purchasing coalitions in the US that use risk adjustment. A third alternative, depicted in Figure 2 which we call modality B, is that the consumer pays the total contribution, i.e. solidarity contribution and premium contribution, to the health plan and that the health plan transfers the solidarity contributions to the sponsor. To reduce the actual flows of money, each health plan and the sponsor net the difference of all the solidarity contributions and premium subsidies for all members of a health plan. This way of organizing the payment flows in a risk adjustment system is being applied in the mandatory sickness fund insurance in Germany and Switzerland and in the voluntary health insurance in Ireland. This modality of organizing the payment flows was also proposed by the White House Task Force on Health Risk Pooling (1993). In Germany the contribution is a certain percentage of the consumer's income. The sponsor requires this percentage to be the same for all members of a sickness fund, but allows it to differ across sickness funds. In Switzerland and Ireland the contribution must be community-rated per health plan (in Switzerland: per region).

Figure 2

As the figures suggest, the direct payment from the consumer to the health plan in Modality A is considerably less than in Modality B. Hence, cost savings by health plans will have a much larger proportional effect on the level of direct payments in Modality A than in Modality B. Both the difference in proportional change and in absolute level of direct payments may result in different responses by consumers [Buchmueller and Feldstein (1997)].

2.2. Subsidy formula

The formula to calculate the risk-adjusted premium subsidies and solidarity contributions can in principle be independent of how the actual payment flows are organized. In practice, however, there is often a relation. Assume, for example, that age is the only risk adjuster. In countries that use modality A (US, the Netherlands, Israel) the health plans receive an age-related subsidy for each consumer, while in countries that use modality B (Germany, Switzerland and Ireland) only health plans with an overrepresentation of elderly receive a subsidy and only health plans with an underrepresentation of elderly pay a risk-adjusted solidarity contribution.

In this Chapter, for convenience, we assume that risk solidarity is fully reflected in the risk-adjusted premium subsidy, and that the solidarity contribution is not risk adjusted. Broadly speaking, this assumption is not restrictive and sacrifices little generality.[19]

For the calculation of the risk-adjusted premium subsidies a central question is: on what costs should the subsidies be based? We shall call these costs the *acceptable costs*. Acceptable costs can be conceptualized as those generated in delivering a "specified basic benefit package" containing only medically necessary and cost-effective care. In principle, the cost of hospitalizations could be excluded when only day surgery is medically indicated; as could the cost of psychiatric care when care by a psychologist is appropriate. Because the cost level of such a benefits package is hard to determine, in practice subsidies are based on observed expenses rather than needs-based costs. This is true of social health insurance programs such as Medicare in the US or the sickness fund systems in Germany, Israel and the Netherlands.

Figure 3

[19] Assume, for example, that age is the only risk adjuster, and that E_i is the average expenditures in age group i, with E the grand average. Assume that the solidarity contribution is $E-E_i$ and has to be paid only by individuals belonging to an age group i with $E>E_i$; and the risk-adjusted subsidy is E_i-E and is received only by individuals belonging to age group i with $E<E_i$. This situation is identical to the situation that each individual pays a non risk-adjusted solidarity contribution E and receives a risk-adjusted subsidy E_i.

Observed expenses are determined by many factors, not all of which need to be used for calculating the risk-adjusted subsidies. Ideally, subsidies should only be adjusted for those risk factors for which solidarity is desired. Society has to decide for which risk factors, and to what extent, it seeks solidarity. Figure 3 summarizes seven classes of risk factors that explain variations in health spending across individuals. The first three groups are characteristics of individuals: age and sex; health status;[20] and socio-economic factors such as lifestyle, taste, purchasing power, religion, race, ethnicity, and population density. The fourth group includes all provider characteristics, such as practice style and whether there is an oversupply of providers or facilities. Input prices are a characteristic of the region in which the providers are located, and are largely exogenous to the patient or provider. The final two groups are characteristics of the health plan. By market power, we mean to indicate the health plan's ability to negotiate price discounts. Benefit plan features include conventional demand side features such as deductibles, copayments and decisions about covered services, but also include supply side features such as utilization review, various health management strategies, and characteristics of the contracts and financial incentives between plans and providers. Even after controlling for these seven systematic factors that affect costs, considerable variation in spending across individuals will remain, which ex ante is random and will be averaged out by health plans by risk pooling. We recognize that not all of these factors are independent, and indeed some are reasonably thought of as partially endogenous to others (we return to this in section 3.4). We use X to denote the full set of risk factors that predict variations in health spending across individuals.

Should all risk factors X that are observed by the sponsor be used to calculate risk-adjusted subsidies? The answer may vary with the sub-population, context and country. In the US, on the one hand, the widespread practice of experience rating health premiums at the employer level is consistent with the view that health premium subsidies by the sponsor (employer) should reflect just about any information that explains variation in spending. On the other hand, individual premium contributions differ greatly across sponsors, and sponsors differ dramatically in how they calculate their subsidy payments to health plans. In Europe, national solidarity is more prized, and there is greater standardization of benefits and sponsor subsidy formulas within each country. Europe is characterized by narrower ranges of individual premium contributions than in the US.

Despite differences in the specifics, most systems implicitly seek to achieve solidarity along some specified dimensions. We divide the risk factors X into two subsets: those factors for which solidarity is desired, the S-type; and those factors for which solidarity is not desired, the N-type. In most societies age, sex and health status are S-type risk factors. Differences in input prices are also likely to be considered S-type risk factors. It may be argued that differences in costs caused by the other risk factors can be influenced by the insurer or by the insured, and should be reflected in the premium contribution. To the extent that the division into S-type and N-type factors is not clear, society should make an explicit choice. For example, hospitalizations for lung cancer, AIDS, obesity, and skiing accidents are all health-related as well as life-style related risk factors. To the extent that consumers and health plans cannot be held responsible for cost differences or to the extent that society decides that solidarity is desired, the subsidies could be adjusted for these factors.

Assume that E(X) is the best estimate of the expected expenses for a person with risk characteristics X in the next contract period. An estimate of the acceptable cost level A(X), which serves as the basis for setting the sponsor subsidies, could then be E(X) with the values of the N-type risk factors set at an acceptable level (e.g., the acceptable level of the price or supply of health care or the acceptable practice style)[21]. The risk-adjusted premium subsidy could then be a function of A(X), e.g., it could be A(X), or A(X) minus a fixed amount (as in the Netherlands), or a certain percentage of A(X) (as in the US Medicare). The calculation of A(X) will be discussed in section 3.

2.3. Regulation

If health plans were fully free to set their risk-adjusted premiums, the set of rating-factors and the resulting range of premiums could be substantial. For example, the premium for private health insurance in the Netherlands may be related

[20] In this chapter we will use the term health status without going into details concerning either the difference between health status and need, or the various concepts of need, such as normative need, felt need, expressed need and comparative need [Bradshaw (1972)]. For a discussion of the concepts morbidity, need and demand, see e.g., Ashley and McLachlan (1985).

[21] For example, in Belgium the weights of the subsidy formula are estimated based on the relevant risk factors including indicators of the supply of health care facilities. However, when calculating the subsidies the differences in supply, an N-type factor, are ignored [Schokkaert and Van de Voorde (1998)].

to age, gender, family size, region, occupation, length of contract period, individual or group contract, the level of deductible, health status at time of enrollment, health habits (smoking, drinking, exercising) and - via differentiated bonuses for multi-year no-claim - of prior costs. Also in the US the premiums for individual health insurance are substantially risk-rated. Insurers commonly use age, gender, geographic area, tobacco use and family size as risk adjusters to determine standard premiums; and dependent on the applicant's health status insurers may charge premiums up to seven times the standard rates [U.S. General Accounting Office (1996, 1998)]. In a competitive health plan market with unregulated premiums, the maximum premium for full health plan coverage (i.e. without cost-sharing) could be expected to exceed the average premium for the same product by a factor 10 or more, with a minimum premium of around 10 percent of the average[22].

To what extent is a system of risk-adjusted premium subsidies able to reduce such a range of consumer payments? In most countries that have implemented a system of risk-adjusted subsidies in a competitive health plan market, age and gender are used as risk adjusters, sometimes supplemented with an indication of disability (the Netherlands) and institutional and welfare status (US). Region often is a controversial candidate for being a risk adjuster, since it can either reflect input cost variation (usually a solidarity factor) or practice style variation (which many may consider undesirable). Risk adjustment models that use only these variables routinely do a poor job. For example, in a simulation based on a simple premium model and subsidy formula, the range of premium contributions was 14,297 Dutch guilders without any risk adjustment versus 11,571 guilders using age and sex to risk adjust. [Van de Ven et al. (1997, Table 7)]. Using age and gender for risk adjustment reduced the range of total individual payments only by 20 percent.

If the resulting range of individual payments is considered to be too large, the sponsor may combine the system of risk-adjusted subsidies with restrictions related to premium contributions and with a periodic open enrollment for a specified basic health plan coverage. A *periodic open enrollment* requirement implies that during the open enrollment period, for example one month every year, consumers are allowed to change plans and each health plan must accept anyone who wants to join.

Restrictions related to the premium contributions can take several forms: community rating, a ban on certain rating factors (for example health status, genetic information, duration of coverage, or claim experience) or rate-banding (i.e. a minimum and maximum premium contribution)[23]. Community rating implies that a health plan must ask the same premium contribution from each individual, independent of the individual's additional risk characteristics. A variant is adjusted community rating, that is, adjustments in the community rate are allowed for various factors (for example claim experience) with various limits imposed on the extent to which rates, after adjustment, may vary. Rate banding can take several forms: per health plan or nation-wide; and may specify either an absolute or a relative difference between maximum and minimum premium contribution. An extreme form of restriction on premium contributions is that health plans are required to accept the individual's risk-adjusted premium subsidy, which is determined by the sponsor, as the full premium. This is the case in the competitive social health insurance systems in e.g., Colombia, Israel and Russia.

The goal of restrictions related to the variation of premium contributions is to fulfil the solidarity principle by creating *implicit* cross-subsidies from low-risk to high-risk individuals (whereas a system of risk-adjusted subsidies implies *explicit* cross-subsidies). However, restrictions on premium contributions also imply predictable profits on low-risk consumers and predictable losses on high-risk consumers. If the premium contributions must be community-rated and if the premium subsidies depend on age and gender only, the health plans will incur substantial predictable losses on their high-risk members. For example, Van Barneveld et al. (1998, Table 2) show that if a health plan were to use information on prior hospitalizations and prior costs in the three preceding years, it could identify a subgroup of 4 percent of its members whose predicted costs are threefold their average age/gender-adjusted expenses. Another example is that the five percent individuals with the highest health care expenditures in any year can be predicted to have per capita expenditures over (at least) the next four years that are twice their average age/gender-adjusted expenses [Van Vliet and Van de Ven (1992, Table 3)]. Ideally, for each health plan the predictable losses on its high-risk members should be compensated by the predictable profits on its low-risk members. However, this ideal situation may not be achieved

[22] In a simulation based on a simple premium model, the minimum premium, the average premium and the maximum premium were respectively 199; 1,500; and 14,496 Dutch guilders [Van de Ven et al. (1997, Table 3)].

[23] Ideally restrictions related to premium contributions should only relate to the S-type risk factors and not the N-type factors. In practice this may be hard to effectuate, especially when S-type and N-type risk factors are correlated.

because of *selection*, i.e. actions[24] by consumers and health plans to exploit unpriced risk heterogeneity and break pooling arrangements [Newhouse (1996)]. Often the term selection is also used to refer to the outcome of these actions. The literature identifies two forms of selection: adverse selection and cream skimming.[25] Because these forms of selection may differ in the consumers' or health plans' actions as well as in their effects on efficiency and fairness, we will discuss each of them.

2.4. Selection

2.4.1. Adverse selection

Adverse selection is the selection that occurs because high-risk consumers have an incentive to buy more coverage than low-risk consumers within the same premium risk group. A necessary condition for adverse selection to occur is that the consumers themselves know whether they are a high- or low-risk *within* their premium risk group, i.e. consumers must have more information about their future risks than the information health plans use for premium differentiation. As Wilson (1977, p. 167-168) highlighted, this consumer information surplus vis à vis the health plan may be caused either by regulation or by a limitation of the health plans' knowledge. That is, either restrictions on premium rates or asymmetric information between health plans and consumers may result in similar adverse selection problems. In the case of asymmetric information the health plans may know that consumers vary in the level of risk, but they cannot discern who are the high- and low-risk individuals within a premium risk group. Pauly (1984) referred to this as "true adverse selection". In the case of regulatory restrictions on health plans abilities to differentiate premiums, the health plans may know the consumer's level of risk, but are not allowed to use this information to set premiums. [26]

Rothschild and Stiglitz (1976) showed that in a market with asymmetric information a competitive equilibrium may not exist. This would be the case if there are relatively few high-risk individuals, which seems a quite realistic assumption for the health plan market[27]. As a result of adverse selection a competitive health plan market may be unstable. Low-risk individuals will persistently (try to) separate themselves from the high-risk individuals by buying new products that are especially designed to lure them from the more heterogeneous risk pool. Premium for the old products will have to rise as they come to be predominantly bought by high-risk individuals. As the low-risk individuals avoid the generous health plans, these plans may be confronted with a fatal spiral of ever rising premiums. Rothschild and Stiglitz showed that if equilibrium exists, high-risk individuals buy full coverage and low-risk individuals buy incomplete coverage (i.e. a separating equilibrium or an "adverse selection equilibrium"). In their model a pooling equilibrium cannot exist.

The strong predictions of the Rothschild-Stiglitz model have subsequently been softened by Wilson (1977), Schut (1995) and Newhouse (1996), among others, who show that pooling equilibria are at least possible. Wilson (1977) shows that if the losses to low-risk individuals from separating themselves from high-risk individuals are greater than the cross-subsidy implied by a pooled equilibrium, then a pooling equilibrium can result. Schut (1995, Chapter 3) shows that costly risk classification may stabilize a competitive health plan market and may result in a Pareto-type welfare improvement. Newhouse (1996) shows that the presence of sufficiently large contracting costs can result in a pooling equilibrium with the low-risk group at its most preferred point and the high-risk group at its most preferred feasible point.

Empirical simulation results by Marquis (1992) suggest that adverse selection is sufficient to eliminate high-option benefit plans in multiple choice markets if health plans charge a single, experience-rated premium. Similar results are found by Keeler et al. (1998). Cutler and Reber (1998) analyzed the health insurance pricing reform by Harvard University in the mid-1990s. Harvard had historically subsidized the most generous plan quite generously at the margin. Under the new policy, Harvard contributes an equal amount per individual/family to each plan regardless of which plan an employee chooses. The plans' premiums are only differentiated for individual/family. Because of adverse selection, the most generous policy could not be sustained under an equal contribution rule (i.e. without risk adjustment). In three

[24] Not including risk-rated pricing by health plans.
[25] For the relevance of the distinction between these two forms of selection, see e.g., Pauly (1984).
[26] For a discussion of regulation-induced adverse selection see e.g., Newhouse (1984, p. 99), Pauly (1984) and Keeler et al. (1998).
[27] When applying the Rothschild-Stiglitz theory in our case, we have to interpret "high-risk and low-risk individuals" as "high-risk and low-risk individuals *within* their premium-risk-group" (e.g., an age/gender-group). The Rothschild-Stiglitz theory then applies to the submarket for each premium-risk-group [see footnote 5 in Rothschild and Stiglitz (1976)].

years the adverse selection “death spiral” was completed at Harvard. Price et al. (1983) analyzed the instability of the Federal Employees Health Benefits Programme (FEHBP), which offers comprehensive benefits to federal workers and retired employees in the US. All FEHBP-plans are subject to annual open enrollment and the premiums are differentiated only according to single/family (that is community rating by single/family class). Price and Mays (1985) found substantial adverse selection within the FEHBP-market. Price et al. (1983) concluded that the FEHBP's lack of stability raises important questions about the viability of some pro-competition proposals involving multiple-insurer systems.

2.4.2. Cream skimming

C*ream skimming* (or preferred risk selection or cherry picking) is the selection that occurs because health plans prefer low-risk consumers to high-risk consumers within the same premium-risk-group. A necessary condition for cream skimming to occur is that the health plans know that there are high- and low-risk individuals *within* the premium-risk-groups. Such a situation may be caused by regulation or by transaction costs related to (further) premium differentiation. Even if there is an open enrollment requirement cream skimming can take place in several ways. Health plans may actively cream the preferred consumers and dump nonpreferred consumers [Ellis (1998)]. The precise form of the selection that may occur, depends on the additional information that health plans have. We distinguish three situations.

First, if health plans only know that there are high- and low-risk individuals within the premium-risk-groups, but they cannot ex-ante identify who are the high-risk individuals and they also don’t know what the relevant omitted risk factors are, they may structure their coverage such that the plan is unattractive for the high-risk individuals [Newhouse (1996); Glazer and McGuire (forthcoming)]. For example, plans may exclude prescription drugs from coverage or may offer a low-option plan with a high deductible and other cost-sharing. In this way health plans use adverse selection as a tool for cream skimming. They stimulate the different risk groups to reveal themselves. Even if the benefits package and the cost-sharing structure are fully specified, health plans may differentiate their coverage conditions by contracting with different panels of providers. For example, a health plan may contract with a selected panel of providers who work according to strict protocols, or it may apply strict utilization management techniques or contract with managed care firms that do so. Such a health plan is more attractive for the low-risk individuals than for the high-risk individuals within each premium-risk-group. Health plans may also share financial risk with the contracted providers in a way that encourages providers to cream skim. Health plans may also try to attract the low-risks by offering a package deal of health insurance and other forms of insurance or services bought mostly by relatively healthy people, including fitness club memberships.

Second, if health plans know that some omitted risk factors are relevant (e.g., AIDS, disability, prior utilization or hypochondria), but they cannot ex-ante identify the individuals with these characteristics, they may deter the high-risk consumers by selectively not contracting with physicians who have the best reputation of treating patients with such problems. Health plans also could contract with providers who have no interpreters, or whose facilities have no disabled access [Luft (1987)]. They may also select by the design of their supplementary health insurance (no coverage for mental health care, prescription drugs and reconstructive breast surgery) or by putting the brochures of competing health plans on the counter in places where sicker people are likely to be, such as in pharmacies and hospitals.

Third, if health plans can ex-ante identify predictably unprofitable individuals based on certain risk characteristics, they can focus their selection strategy directly on those identifiable individuals, e.g., by providing the high risks with poor quality of care or poor services (such as delayed payments of reimbursement and delayed answers to letters); by not working to coordinate the multiple visits that people with many problems may need; by selective advertising and direct mailing; by contracting with providers who practice in “healthy districts”; by providing the insurance agent with incentives to advise relatively unhealthy persons to buy health insurance from another company; or by a golden handshake for unhealthy members at disenrollment, such as offering an AIDS patient a large sum of money to choose a different plan during the next open enrollment.

2.5. Effects of selection

The primary rationale for regulating a competitive health plan market is to provide financial access to health plan coverage for the high-risk individuals. Because regulation induces selection, we have to understand the effects of selection to evaluate the overall effects of regulation.[28]

As stated above, depending on the relative proportion of high-risk individuals within each premium-risk-group and contracting costs, *adverse selection* may either cause a competitive health plan market to be unstable or it may result in a pooling equilibrium or it may result in a separating equilibrium. In the last case high-risk individuals pay a high premium for generous coverage and low-risk individuals pay a low premium for stingy coverage. So, adverse selection may decrease access to coverage for non-affluent high-risk individuals. The inefficiency that arises in an adverse selection equilibrium is that, depending on the contracting costs, either the low-risks or the high-risks cannot obtain as much coverage as they wish[29]. Another inefficiency arising from adverse selection is the welfare loss due to the potential non-existence of a competitive equilibrium. The continuous exit (bankruptcy) and re-entry of health plans has real social costs.

Even with a periodic open enrollment requirement (to prevent health plans' refusing relatively high-risk individuals) there may be *cream skimming*. First, the larger the predictable profits resulting from cream skimming, the greater the disincentive for health plans to respond to the preferences of high-risk consumers. Health plans may give poor service to the chronically ill and choose not to contract with providers who have the best reputations for treating chronic illnesses. This in turn can discourage physicians and hospitals from acquiring such a reputation. To the extent that a health plan and its contracted providers of care share financial risk, the providers share the incentive to attract profitable patients and to deter patients who generate predictable losses. As Newhouse (1982) highlighted in his famous "mother with an asthmatic child" example, providers of care have subtle tools to encouraging high cost patients to seek care elsewhere, such as keeping the patient in uncertainty about the correct diagnosis, making the patient wait for an appointment, making the patient wait in the office, being discourteous to the patient, or advising chronically ill patients to consult another physician who is "more specialized in treating their disease". Health plans who specialize in care for high-risk patients, have to ask a high premium (because of adverse selection)[30]. So, as a result of selection, high-risk patients may either receive poor care and poor service or pay a very high premium for good care and good service. If the regulation implies a nation-wide maximum premium instead of a maximum per health plan, health plans that experience adverse selection cannot raise their premium and will go bankrupt. In that case, it is suicidal for a plan to become known for providing the best care for chronically ill, because it will be flooded by individuals who predictably generate more costs than revenues.

Second, the larger the predictable profits resulting from cream skimming, the greater the chance that cream skimming will be more profitable than improving efficiency. At least in the short run, when a health plan has limited resources available to invest in cost-reducing activities, it may prefer to invest in cream skimming rather than in improving efficiency. In the long run, improving efficiency may be rewarding, independent of the level of cream skimming, as long as these improvements are perceived as desirable by consumers.[31] Efficient health plans who do not cream skim applicants, may lose market share to inefficient health plans who do, resulting in a welfare loss to society.

Third, while an individual health plan can gain by cream skimming, for society as a whole, cream skimming gains nothing. Thus, any resources used for cream skimming represent a welfare loss.[32]

In sum, regulations that are intended to increase access to coverage for high-risk individuals may instead induce selection efforts with the following unintended effects (see Table 1): problems with financial access to coverage for high-risk individuals, reductions of the quality of certain kinds of care, or reduction of allocative efficiency and efficiency in the

[28] Another effect of regulation is that it limits health plans in designing and pricing their products (e.g., managed care and no-claim bonuses) such that they reduce undesirable moral hazard.

[29] Of course, the *desired* level of health plan coverage depends on the tradeoff between moral hazard and risk aversion [Zeckhauser, (1970); Manning and Marquis (1996)].

[30] In the short run, a small health plan which specializes in care for high-risk patients may be confronted with financial problems if, *after* it has determined its premium for the next contract period, it is flooded by a group of high-risk members.

[31] Cost-reducing efficiency gains need not always be desired by consumers that receive subsidized insurance premiums. It may fall upon the sponsor to decide what the acceptable costs are.

[32] Resources used by health plans for product innovation or for designing contracts which provide consumers an incentive to become/remain in good health, but which may also attract low-risk individuals, are not considered a welfare loss [Beck and Zweifel (1998)].

production of care. So, given a system of imperfectly risk-adjusted subsidies, there is *a tradeoff between access to coverage and the adverse effects of selection.* A relevant question therefore is: How can we prevent selection?

Table 1

2.6. How can we prevent selection?

Theoretically, the best strategy to reduce selection is *good risk adjustment* (see section 3), so that the heterogeneity of the subsidy-risk-groups is small and the expected cost of cream skimming exceeds its expected profitability. The more homogeneous costs are within a rate category the harder it will be for health plans to attract only enrollees whose average expected profit is high.[33] Whether feasible levels of risk adjustment still allow serious adverse selection remains an empirical question [see e.g., Pauly (1986)]. In case of perfect risk adjustment, there is no selection.

As perfect risk adjustment is still a long way off, a second strategy to reduce selection is *risk sharing* between the sponsor and the health plan, which we discuss in section 4. However, risk sharing reduces a health plan's incentive for efficiency, causing a tradeoff between selection and efficiency.

A third strategy to reduce selection is to allow health plans to *risk rate the consumer's premium contribution* within a certain range. Consequently, any information surplus the health plans might have over the sponsor would be focused on premium differences rather than on cream skimming. This could potentially worsen access for the high-risk individuals, yielding a tradeoff between access and selection. If health plans are required to identify any risk factors they use for premium differentiation, the sponsor could try to include these risk factors in the subsidy formula in subsequent years, thereby reducing the potential for cream skimming. Potentially, market-driven improvements of the risk adjustment mechanism may be more effective and more workable than research-driven improvements.

Several *additional measures* may be adopted to reduce selection. One straightforward way to prevent an extreme form of adverse selection - that is, one in which low-risk individuals do not buy health plan coverage at all and thereby do not cross-subsidize the high-risk individuals - is to mandate that everyone must buy some minimum basic health plan coverage. Mandating a minimum health plan reduces but does not eliminate the possibility that health plans may differentiate their insurance plans to try to enroll profitable individuals. Forbidding selective contracting, such as by imposing an "any-willing-provider" mandate is a related tactic. Given the many subtle ways health plans can differentiate the coverage of their benefits package, this type of regulation may be hard to enforce. However, even if a sponsor could successfully implement mandatory health plan membership with uniform conditions, it could have several adverse effects[34]. First, it would impede health plans from selectively contracting with only cost-effective providers. This reduces the potential for managed care activities by the health plans, implying a loss of efficiency in production. Second, a "one-size-fits-all" plan reduces the consumer's choice and yields a welfare loss[35] because it reduces the health plans' responsiveness to consumer preferences. Third, a standardized plan reduces the health plan's initiatives to design insurance contracts that reduce undesired moral hazard. Fourth depending on the generosity of the fully standardized benefits package, a mandatory health plan membership may increase moral hazard problems.[36]

A second and closely related measure for reducing selection incentives is to "carve out" or separately cover services on which health plans may potentially have the greatest incentive to select. Pharmaceuticals, mental health treatment, and dental care are frequently not included in the standard benefit package, but are either not covered or covered separately. The classic rationale is concern about demand side moral hazard response since these services appear to be more price responsive to insurance coverage [Morrisey (1992)]. More recently Frank et al. (1997) and Ettner et al. (1998) have

[33] Although a refinement of the subsidy formula *on average* lowers the profits of cream skimming, for some individuals it might *increase* the profits [*see e.g.*, Beebe et al. (1985)]. Therefore a detailed exploration of the distribution of the potential profits and losses per individual insured may be necessary.

[34] In contrast to these adverse effects, a certain degree of standardization may have the advantage of making the market more transparent and reducing the consumers' search costs.

[35] For an estimate of this welfare loss, see Keeler et al. (1998).

[36] In addition, even if the implementation of mandatory health plan membership with uniform conditions could successfully prevent adverse selection in a competitive health plan market, health plans would still be left with other tools for cream skimming, such as tie-in sales, selective advertising, design of supplemental health insurance, providing poor services to high-risk individuals, selective advise of insurance agents and a golden handshake.

examined the rationale behind carving out these services, which includes the fact that these services are more predictable, and hence more prone to selection activities.

A third additional measure to reduce selection might be to increase plan level entry or exit barriers. The qualification or certification of health plan contracts by the sponsor or by an independent organization will make it more expensive for plans to enter so as to cream skim, or exit so as to avoid adverse selection. Sponsor subsidies can be earmarked for the purchase of qualified or certified health plan contracts only. The requirements for qualification of health plan contracts may relate to the design of the benefit package, the copayment structure, the quality of the contracted specialty-mix, the forms of risk sharing between the health plan and the contracted providers, the location and accessibility of the contracted facilities, etc. The pricing and selling of qualified health plans should not be tied-in with other products and services.

Fourth, regulations of the enrollment procedure may influence selection activities. Enthoven (1978, 1986) proposed that there be no direct interaction between a health plan's sales representative and a potential member in the enrollment process. The potential members should deal with an independent agency (or the sponsor itself) that notifies the health plans of those who have enrolled for the coming contract-period. Every family would receive a booklet, published by the administrative agency, containing meaningful, useful information on the features and merits of the presented alternatives. Furthermore, the contract period should not be too short. The shorter the contract period, the higher is the proportion of predictable episodes of costly illnesses (predictable by both the health plans and the consumers) during the next contract period(s). An example is the potential dumping of some patients at high risk of death [Newhouse (1986)]. Switching plans to take advantage of better pregnancy and birth benefits is another important example. The short (one month) lock-in period for a Medicare insured who chooses an HMO provides many opportunities for selection. The one-year lock-in period as applied in the Dutch sickness fund market may be a good compromise between sufficient consumer choice and not too much selection. Pauly (1988) proposed requiring consumers to choose their health plan option a long time before the renewal date of the contract. This lowers the predictability of future costs during the new contract period and thereby reduces the potential profits of selection.

Fifth, improved consumer information may mitigate selection, particularly monitoring and publicizing of information on plan quality. Luft (1982) suggested monitoring systems in which people who change plans are asked about any problems they experienced and whether they felt pushed out. Such information and data from more broadly targeted consumer satisfaction surveys could be very worthwhile for the consumer. The sponsor also could examine the health care needs and costs of those consumers who switch plans as a way of monitoring health plans' (and the contracted providers') behavior. In theory, the sponsor could raise the cost of cream skimming by dissemination such information to consumers. In practice, this methodology probably has even further to go than risk adjustment or risk sharing.

Sixth, ethical codes for health plans might be designed to reduce incentives to select. Codes could be developed either by the government or by professional organizations; violation of these codes could be a punishable offence. The ethical codes could relate to things such as the quality of the contracted providers, procedures for making and handling complaints, selective advertising, golden hand-shake, etc.

Seventh, the sponsor will need to evaluate and periodically adjust and improve the risk-adjusted subsidy formula over time. Risk adjustment should not be done once and left alone. The sponsor will need to update the risk adjustment formula in light of technological change or behavioral responses to risk adjustment by the health plans and by consumers. The credible announcement by the sponsor of its intention to periodically improve on the accuracy of its risk adjustment methodology will reduce the expected profitability of certain cream skimming activities, and may lessen their use.

The extent of the success of these measures to prevent selection largely depends on the size of the predictable profits and losses that result from the regulation, as well as on the costs of selection, including the cost to a health plan of losing its good reputation.

3. Risk adjustment models

In this section we concentrate on the theoretically most preferred strategy to reduce selection, i.e. risk adjustment. We examine the specific risk factors and models that can be used for calculating the best estimate of acceptable costs. By *acceptable costs* we mean the cost of the set of services and intensity of treatment that the sponsor has chosen to

subsidize, as defined in section 2.2. We begin with a discussion of criteria that can be used for assessing risk adjustment models, and apply these criteria to issues related to designing, evaluating, and choosing a model. Specifically, we consider: criteria for selecting a risk adjustment model; choice of prediction period; choice of explanatory variables to use for risk adjustment; selection of a functional form; and use of summary statistics to assess and compare alternatives. The section ends with a review of selected state-of-the-art risk adjustment models that are compared in terms of their ability to achieve the objectives set out at the beginning of this section.

3.1. Criteria for choosing among risk adjustment models

A number of very useful surveys of risk adjustment models have proposed criteria for comparing different risk adjustment models [Thomas et al. (1983); Newhouse (1986); Epstein and Cummella (1988); Van de Ven and Van Vliet (1992); U.S. General Accounting Office (1994), Ingber (1998)]. Although more than a dozen criteria can be listed, they can usefully be grouped into three broad criteria, which may be mutually related:

- Appropriateness of incentives
- Fairness
- Feasibility

In addition to the "appropriateness of incentives", an efficiency concept, and "fairness", which have already been emphasized, we see here the new concept of "feasibility." The feasibility of risk adjustment models imposes constraints on the key tradeoff between efficiency and fairness discussed in previous sections. Although a perfect risk adjustment model might be able to eliminate this tradeoff, such a model might not be feasible to implement.

3.1.1. Appropriateness of incentives

Correcting for selection and moral hazard problems are the primary reasons for implementing risk adjustment. Thus, the most important criterion for evaluating risk adjustment models is by the extent to which they create *appropriate incentives*. There are many possible distortions or undesirable responses to risk adjustment , in particular when combined with restrictions related to the premium contributions and with open enrollment. Since it is an area of keen interest and research by economists, Table 2 provides an extensive list of the ways that provider and health plan behaviour may respond to incentives created by the risk adjustment and the various regulations.

Table 2

It is beyond the scope of this chapter to fully review the literature on each of these topics. The literature has traditionally focused on how benefit design and premiums influence plan selection by enrollees who differ in expected health [Morrisey (1992); Jensen and Morrisey (1990)]. Differences in expected costs that result from cost sharing differences should in most cases be taken into account when developing and implementing risk adjustment models [Van de Ven and Van Vliet (1995)].

Plans can use a range of strategies for attracting profitable enrollees and avoiding unprofitable ones, such as by denying coverage, exclusions for preexisting conditions, and selective enrollment or disenrollment counseling. Many of these strategies are regulated or prohibited in some, but not all, countries and settings. These strategies should probably be addressed through regulation rather than asking risk adjustment models to solve all of the problems of creating an effective health care market. (see section 2.6).

In the presence of government regulations prohibiting explicit selection, health plans have incentives to manipulate the specific services that they offer to enrollees. This topic has been the focus of a recent flurry of research, perhaps reflecting growing concerns about its potential importance. Ellis (1998) develops a framework in which health plans have incentives to oversupply services to profitable patients ("cream skim") and undersupply ("skimp on") or "dump" (avoid treating) patients that are unprofitable. Improved risk adjustment reduces the incentive for plans to engage in these activities, but also changes the particular enrollees that plans will compete to attract. For example, increasing payments for individuals expected to cost more than the average can result in plans competing to attract such individuals, a reversal of the incentives with unadjusted capitation payments.

Although premium subsidies that are fully adjusted for the consumer's health status make selection unimportant, these payments may be criticized because they create inappropriate incentives for health-improving activities. One could argue that a health plan that improves its members' health status by good quality care and effective prevention is penalized by lower future revenues [McClure (1984); Luft (1996)]. A counter argument, however, is that improved health status not only reduces future revenues, but also future expected expenditures. Furthermore, if a health plan effectively reduces the incidence of lung cancer or heart diseases, it fully benefits from not having the high first-year expenses related to these diseases.[37] In addition the plan fully benefits from not having expenses related to preventable transitory health problems for which the subsidies are not adjusted (e.g., fever and flu). Nevertheless, it is true that a health plan bears the full costs of health-improving activities and preventive services such as smoking cessation, weight loss, and nutritional guidance, while it may lack a part of the future returns. In other words, from the point of view of the health plan health adjustment may reduce the cost-effectiveness of some prevention programs. Whether in practice these incentives override the professional ethics of the providers and the consumer preferences, remains an empirical question.

McClure (1984) suggested the following two solutions. The first is to make bonus adjustments based on change in health status over time. With care and ingenuity, it may be possible to devise subsidies that reward health improvement but that cannot be gamed by the plans. Secondly, McClure suggested making public to beneficiaries any change in overall health status levels in each health plan. So beneficiaries might shop for health plans on the basic of health status improvement figures. Plans would thus gain a reward for improving health status by attracting new enrollees. A third solution is to provide health plans with earmarked payments for effective prevention programs.

Several studies have discussed the incentive for health plans to distort information reported to the sponsor if that information is used for payment purposes [e.g., Epstein and Cumella (1988)]. This may occur either with diagnosis- or survey-based risk adjustment. Carter, Newhouse and Relles (1990) examined changes in diagnostic coding in the United States in response to the Medicare Program's payment system based on Diagnosis Related Groups (DRGs), which they termed "DRG Creep". They also suggest that such changes in diagnostic coding appear to be one-time. It seems plausible that similar responses might occur from risk-adjusted capitated payments, but we are not aware of any studies documenting this result empirically to date.

The predictive accuracy of different models is by far the most common criterion on which risk adjustment models are compared. Yet the goal of risk adjustment is not accuracy per se, but rather improved incentives and fairness. Using prior information that is known to the individual or plan to adjust payments is important because it should lessen the danger of cream skimming or dumping. Specific measures of predictive power are discussed below, along with a consideration of whether it is individual or group resources that should be predicted.

Although greater predictive power is generally desirable, it is important to emphasize that higher predictive power is not necessarily preferred to less. For example, actual expenditures are perfectly correlated with actual expenditures, and are an excellent "predictor" of the health care use in that same year. Yet such fee-for-service reimbursement is a very imperfect basis for payment since it creates undesirable disincentives for efficiency, and "costs" are difficult to measure and monitor. Similarly, models that base their predictions upon the type of service provided, the use of specific procedures, or concurrent year diagnoses, can be more accurate, but may be create inappropriate incentives.

Finally one may argue that mortality as a risk adjuster provides health plans with inappropriate incentives ("mortal hazard").

3.1.2. Fairness

We discussed fairness within the framework of the solidarity principle of sections 1 and 2. While the fairness of the method of collecting premiums and calculating risk adjustment subsidies has been the topic of considerable discussion in many European countries, it has received considerably less attention in the United States. For example, the *fairness* of the risk adjustment model does not enter explicitly into the list of criteria used to compare across different models in the reviews of Epstein and Cumella (1988), the U.S. Government Accounting Office (1994), or Ingber (1998).

[37] This argument does not hold in case of retrospective risk adjustment (see section 3.2.2.).

Decisions about fairness and about what risk factors should be labeled an S-type or N-type factor, reflect value judgments that differ across countries and among individuals. There appears to be a consensus that factors that reflect purely tastes (e.g., religion or a preference for cost-ineffective care) may have predictive power but do not belong in a risk adjustment model based on commonly held fairness principles. Lifestyle is a more problematic risk factor. On the hand one could argue that health care expenditures that are purely related to smoking or sexual behavior should not be subsidized because these expenses can be influenced by the individual. On the other hand, many people will argue that these expenditures should be subsidized because it is unfair if people with lung cancer or AIDS cannot receive an appropriate medical treatment.[38]

Another discussible factor is average distance between patients and providers or density. Should the premium subsidies be lower for geographically dispersed regions with poor access to health services? In the United States, the Medicare program's formula for reimbursing HMOs (in 1998) fully reflects county level geographic variation in average health costs, but it is not clear that it should do so.[39] Some of the geographic variation in health costs is due to differences in cost of living between different regions. Many people consider it fair that risk-adjusted subsidies for persons living in high wage cost regions, where medical care is more expensive, should be higher than those for people living in low wage cost regions. This argument holds in particular if the solidarity contribution in the high wage cost regions is higher than in the low wage cost regions. But if the variation is due to practice style variation, taste differences, over- or undersupply, or differences in access, geographic adjustment may be viewed as unfair. The same argument may hold for factors that primarily reflect differences in access, such as race, minority group or ability to pay (income). By not adjusting the subsidy for these access indicators, individuals with poor access will either become preferred risks, which may increase their access, or they will pay a lower premium contribution.

A different type of equity argument is that individuals who are sicker should have risk-adjusted subsidies that are higher than for those who are less sick. This implies that evidence of a new disease or chronic condition for a person should never result in a reduction in the risk-adjusted subsidy for that individual if there exists a cost-effective medical treatment for the person's health problem. This equity argument (monotonicity) does not always hold in empirically derived risk adjustment models. For example, in the empirical risk adjustment models described in Ellis et al. (1996b), in many specifications it was found that among US Medicare enrollees, individuals classified as having dementia (e.g., Alzheimer's disease) have lower predicted medical costs than persons with otherwise identical demographic and diagnostic information. If this reflects underutilization, it seems unfair to reduce payments for this group, even if it is predictive of lower costs. Similarly, Ash et al. (1998) find that in some samples those with profound and severe mental retardation have lower predicted costs than those with mild retardation. If this lower utilization reflects underutilization, one may argue that it should not be reflected in the subsidies for fairness reasons. In this way the underserved become the preferred risks, which may reduce their underutilization.

As we suggested in the preceding subsection, a risk adjustment system will often be considered fairer if it predicts a larger proportion of the variation in health spending. If health plans are fully compensated for the higher expected costs of enrollees with chronic conditions, then it is more likely that they will enroll them, thereby increasing the access of these high cost people. In addition, health plans will bear less risk. Yet as the above examples highlight, improved accuracy that comes from using information for which solidarity is not desired or from risk factors indicating poor access or underutilization may worsen rather than improve fairness.

3.1.3. Feasibility

Administrative *feasibility*, closely related to the criteria of obtainability discussed in Van de Ven and van Vliet (1992), is the requirement that the measures are feasible to obtain for all potential enrollees without undue expenditures of time or money. Information that is routinely collected, standardized and comparable across different health plans, and measures that are easily validated have greater feasibility than measures that require separate data collection, validation and processing.

[38] An alternative is to let the solidarity contribution partly depend on lifestyle factors. E.g., a surcharge on tobacco could go to the sponsor.
[39] The U.S. Balanced Budget Act of 1997 seeks to reduce differences among county level averages used for risk adjustment in the Medicare program.

A further dimension of feasibility is that large, representative samples exist on which risk adjustment models can be developed and parameterized prior to implementation, or used for recalibrating subsequent to adoption. This weakness is particularly serious for survey-based predictors. Another dimension of feasibility is length of the time lag required between the collection of data and its feasible use for payments. Long lags between the date when a health service is provided and the date on which a claim is submitted and processed, can constrain the feasibility of diagnosis- or other claims-based risk adjusters.

Risk adjustment will be feasible only if it is accepted by consumers, providers, health plans, and sponsors. Although a considerable amount of academic research has gone into improving the predictive power and incentives of risk adjustment, relatively little has been published on making risk adjustment acceptable to all parties involved.

One dimension of acceptability is that a risk adjustment model should not compromise the right to privacy of consumers and providers [Epstein and Cumella (1988); and Van de Ven and Van Vliet (1992)]. For example, a risk adjustment approach that requires individuals or providers to identify specific individuals who are HIV positive or who suffer from mental illness may be unacceptable to consumers, regardless of other merits.

Race, ethnic background, and religion are examples of demographic variables that may not be acceptable for risk adjustment primarily due to concerns about fairness. Paying more to a plan during the year in which an enrollee dies may be an actuarially good way to recompense it for the known high costs incurred in the last months of life. However many are repelled by the idea of paying more to health plans because their mortality rates are higher.

Clinical credibility is another dimension of acceptability, since doctors and clinically trained health administrators are important decision-makers. Regardless of whether it affects the predictive accuracy of the risk adjustment model, if clinicians see large differences in payments based on apparently trivial classification differences, then this will undermine acceptability to clinicians.

One last group for whom acceptability is central is actuaries, who typically work for sponsors or health plans and traditionally calculate premiums and provider payments based on demographics and prior experience measures. An important criterion for them is that risk adjustment models are actuarially fair. In the United States, actuaries have been slow to accept health-based risk adjustment, despite its greater accuracy.

3.2. Preliminary issues in designing or implementing risk adjustment

3.2.1. Individual versus contract level risk adjustment

As stated earlier, we take it as given that it is desirable to calculate health-based payments at the level of individuals rather than contracts, such as families or employers. Actuaries in the United States and elsewhere often focus attention on calculating expected payments at the contract level, with the employee and all dependents counting as one unit of analysis. However this approach, focusing solely on the number of people and their relationship to the enrollee without regard to the age and sex breakdown, ignores obviously important information. According to our approach, the expected payment at the contract level (family or employer) can be calculated as the sum of the expected payments for the covered individuals. Although we understand the actuaries' argument that in a competitive market an insurer has to break even on each insurance contract and not on each insured person, the advantage of our approach is that when one individual (e.g., an HIV patient) goes from contract unit A to B, we can easily recalculate the expected payments at the contract level.

3.2.2. Prospective versus retrospective use of risk adjustment information

In developing or implementing risk adjustment, important choices must be made about how information will be used. One alternative is that payments are calculated prospectively, at the beginning of the prediction period using only prior information. A second alternative is to calculate payments retrospectively, at the end of the period. Retrospective payments can reflect information that becomes known during the period being predicted. As Ellis and McGuire (1986) and Newhouse (1986) have highlighted, these two extremes are not the only ones possible: one can also make payments that are a mixture of the two. We focus here on the two pure cases, and defer to section 4 the discussion of risk sharing arrangements that are implied by taking combinations of the two.

The percent of the variance in health spending at the individual level that is predicted using a retrospective framework is considerably greater than what can be predicted prospectively. However, a retrospective framework may not be preferable in practice. While there are estimates of the maximum potential variance predictable by prospective risk adjustment models (see section 3.2.6.), we do not have a standard for how much variance a good retrospective adjuster should predict [Newhouse et al. (1997)]. The incentive and fairness properties of retrospective adjusters are not inherently superior, and the feasibility of using retrospective models is probably worse. Dunn et al. (1996) compared the predictive accuracy of prospective and retrospective frameworks on groups of enrollees and found surprisingly small differences in predictive power for groups when the samples were reasonably large. Ellis et al. (1996b) and Ash et al. (1998) likewise find that prospective models do nearly as well as retrospective models when nonrandom groups of individuals are formed using only prior-year information. Chapman (1997) finds a greater advantage of retrospective models over prospective models in his plan level analysis, but he focuses primarily on group level predictions rather than individual level predictions.

Conceptually, an argument for preferring the use of prospective information for risk adjustment is that only prospective information is potentially known to health plans and individuals at the time that they are making enrollment decisions, and hence used for risk selection. Prospective models attach relatively more weight to information related to chronic conditions that persist over time, while retrospective models attach more weight to information that signals the presence of acute problems. If two persons are ex ante observationally identical, but ex post only one of them turns out to have a heart attack, then under a wide range of assumption, it should not matter for incentives on the plan whether they are compensated ex post for the actual cost of the one getting the heart attack, or ex ante for the expected cost of the likelihood that one of the two will have a heart attack. Newhouse et al. (1997) highlight that explaining truly random events is unimportant when the risk is averaged over many conditions and many individuals. On the other hand, if there is moral hazard on the probability of having the heart attack, or discretion in the treatment of and recording the acute diagnosis of heart attack, then the two systems are not the same. In the US, for example, many believe that there is too little prevention and too much treatment. In such an environment, paying prospectively rather than retrospectively will create superior incentives to avoid and not over-diagnose heart attacks. This moral hazard problem is potentially quite important for the many health conditions for which treatment or prevention activities are discretionary.[40]

Although one may give a high weight to the above argument, it ignores that a retrospective framework protects health plans against adverse selection by individuals with a diagnosis that yields high costs in the period (e.g., a year) in which the diagnosis is set and from which moment it can be used as a risk adjuster. If this argument is relevant, which still is an empirical question, sponsors may consider to extend a prospective risk adjustment model with selected one-year retrospective elements.

Prospective models tend to be more feasible than retrospective models. As a practical consideration, prospective frameworks have the advantage that the information is available sooner, and health plans have more predictable revenues at the beginning of each prediction period. This predictability is attractive both for plans and for sponsors.[41] A second practical consideration is administrative feasibility of available data. Developing a retrospective model has the advantage of only requiring data from a single period, versus two for prospective modeling. Implementing each model imposes similar data collection burdens.

Although the arguments are not all unambiguously in favor of a prospective setting, our interpretation weights the arguments in favor of a prospective framework as relatively more important. Therefore, we focus our attention in this chapter primarily on prospective risk adjustment models. For clarity of presentation, we describe the various models as if a prospective setting is the only intended use. We return at the end of this section to compare various prospective and retrospective models, and in Section 4 we compare various risk sharing strategies that share much in common with retrospective adjustment models.

3.2.3. Functional form

There is a considerable literature in statistics, econometrics, and health economics that examines and assesses alternative functional forms for estimating models of health spending. Although these models often include many

[40] A contrary view is that prospective payment may overpay for persons with high-blood pressure who don't use any medicine.
[41] On the other hand, the predictability of a health plan's margin is higher under a retrospective model than a prospective model.

N-factors, and not just S-factors that policy makers and researchers are interested in using for potential risk adjustment models, this literature has an important bearing on the selection of models. The classic article in this literature is that of Duan et al. (1983), which developed the so called "two part model" of health spending. This model decomposes the expected level of spending (Y) given a vector of explanatory variables (X) into the two parts using the identity:

$$E(Y) = Pr(Y > 0 \mid X)\ E(Y \mid Y > 0, X)$$

Several specifications have been used for each part of this model, including Probit, logit, and linear probability models for the first part, and linear, log-linear and square root models for the second part, which is conditioned only on observations that are strictly positive.[42] Both the use of two part models and nonlinear transformations of the second part are used to improve consistency of the ordinary least squares (OLS) model given the highly heteroskedastic errors. Conventionally, both parts of the two part model are estimated independently and a smearing transformation [see Duan et al. 1983)] is used to generate unbiased estimates of the second part of the model in the common situation in which nonlinear transformations of the dependent variable are used. The classic article using this approach is Manning et al. (1987).

Since this issue is already examined at length in this Handbook in chapter XXX by Andrew Jones (which examines econometric issues), we highlight here only two observations based on the recent literature relevant for applications of risk adjustment in practice. The first observation, made by Mullahy (1998), is that for the two part models to yield unbiased estimates of both partial effects and conditional means, it is important that the error structure strictly satisfy the homoskedastic error assumption, or else a nonlinear smearing correction can lead to seriously biased estimates. This point is reinforced by the companion article by Manning (1998) which demonstrates that predicted means can be seriously underpredicted (e.g., 20%) if heteroskedasticity is not taken into account. Manning makes the important point that the use of the simple transformation log(Y+1), motivated by its convenience, has very poor statistical properties for use in risk adjustment.

The second point is that rather than using nonlinear, two-part models of health spending, the problem with health spending having a thick upper tail can be dealt with by using extremely large samples, and correcting standard errors for heteroskedasticity using the Huber/White formula. Mullahy (1998) notes in a footnote that when sample sizes are large, using simple nonparametric techniques such as cell means or linear regressions may be sufficient, an argument that we find convincing. Monte Carlo simulations presented in Ellis and Azzone (1998) suggest that the attractiveness of simple linear models relative to two part models increases as the predictive power of the risk adjustment models increase. With only a few exceptions, the major risk adjustment models discussed below have used simple linear models, for which there is no retransformation problem. Another argument for the use of simple linear models is to stay as close as possible to the cell-based approach, i.e. the calculation of the average expenditures per risk group, which is mostly used by sponsors for risk adjustment and by health plans for premium rating. For the remainder of this chapter, we focus on simple linear models, comparing them to nonlinear models only to make a point about the effect of nonlinear transformations on measures of predictive power.

3.2.4. Adjustments for partial years of eligibility

It is very common for people to be eligible for health coverage for fractions of a year. This happens automatically with births and deaths, and may also occur due to enrollment or disenrollment. This presents a problem for risk adjustment models both in terms of efficient estimation and in terms of prediction.

It is clearly undesirable to simply exclude those with partial years of eligibility when the goal is unbiased prediction, since partial year eligibles tend to be systematically different from average. Simply including observations of those with partial years of spending without any recognition of the partial year eligibility is also undesirable, since the resulting models will tend to underpredict spending if the model is used to make predictions that are used for partial year rather than full year payments.

[42] For useful references and discussion of this literature see Mullahy, (1998). Also relevant are the debates of the 1980s, most notably that of Hay and Olsen (1984), which examined the desirability of estimating two part models while assuming that $E[Y|Y>0, X]$ can be estimated consistently using only observations where $Y > 0$.

Consider the following example involving only two persons. Suppose person A is eligible for all 12 months and costs $6,000, while person B is eligible for only 6 months, but costs $12,000. Total spending on these two persons is $18,000, and total eligible months are 18, so the correct monthly average is $1,000 per month, or 12,000 per year.

Two corrections for partial year eligibility have been made in the literature, one focusing on unbiasedness, the other focusing on maximizing statistical efficiency. Ellis and Ash (1995) argue that spending for partial year eligibles should be annualized, and then each observation should be weighted by the fraction of the year that the person is eligible. Hence in the above example, person B has $24,000 in annualized expenditures, with weight .5. The weighted average is then (6,000*1 + 24,000*.5) /1.5 = $12,000, which gives the correct annual average. An alternative approach is developed in Hornbrook et al. (1998), who assume that person A reflects 12 draws of monthly spending, while person B reflects only six draws. If the monthly draws are independent and homoskedastic, then efficient weighting reverts to the formula used by Ellis and Ash. However if the monthly draws are correlated, (which empirically they are), then the efficient weights are to place relatively less weight on person A relative to person B than the ratio 2:1. Alternatively, once heteroskedasticity rather than correlation is modeled, empirically it is generally true that the monthly draws for people with shorter eligibility have a higher monthly variance than those with a full year of information. Hence efficient weighting would place relatively less weight on person B. It is easy to show that predictions based on either of these two weighted least squares models, in general, will generate biased estimates of the sample means. Whether they are more accurate predictors empirically does not yet appear to have been answered.

3.2.5. Determinants of R^2

A common measure of the predictive power of different risk adjustment models, but by no means the only one, is the conventional R^2, which measures the proportion of the variance in individual expenditures that is explained by a set of risk adjusters. Nearly all empirical studies on risk adjustment present R^2-values. Ideally, in order to prevent overfitting R^2-values should be reported which are based on out-of-sample predictions. In that case Efron's (1978) R^2 should be used. Some studies have dealt with the question what the maximum R^2 is that can be achieved by a set of prospective risk adjusters (see section 3.2.6). For the interpretation of R^2-values presented in the literature it is important that the R^2 (as well as the total variation) may depend on: (1) the type of services under analysis; (2) the (sub)population under analysis; (3) the variation in explanatory factors; (4) the level of medical technology; (5) the year of the data analyzed, and (6) the length of the time period being predicted. We discuss each of these determinants, which may be mutually related.

The relation between R^2 and *type of service* can be illustrated as follows. Newhouse et al. (1989) found an R^2 of 0.05 for inpatient care and an R^2 of 0.25 for outpatient care, using the same comprehensive set of risk adjusters for the same population (14-64 years old); the R^2 for total acute care was 0.09. Wouters (1991) also found much higher R^2-values for outpatient expenditures than for inpatient expenditures, using the same set of adjusters. In addition, she found that among the various types of outpatient services there is a wide variability in out-of-sample prediction R^2-values, using the same set of adjusters. Drugs ranks first (R^2= 0.40), followed by visits, diagnostics, procedures, and surgery (R^2= 0.005). Van Barneveld et al. (1997) analyzed expenses for several forms of expensive long-term care, like institutional care for mentally handicapped persons, nursing home care and institutional psychiatric care. Using 2 year prior costs as a risk adjuster they found an R^2 of 0.56. This figure is much higher than the comparable R^2-values for acute care, which typically are below 0.15.

The relation between R^2 and *subpopulation* can be illustrated by the results of Kronick et al. (1995). Analyzing US Medicaid claims they concluded that expenditures are much more predictable among persons with Medicaid entitlement based upon disabilities than for other populations. Using prior year expenditures as a risk adjuster they found R^2–values on four different data sets ranging from 0.29 to 0.51. In explaining these relatively high R^2 Kronick et al. suggest that among persons with disabilities a much greater portion of resource utilization results from chronic problems and their complications which persist from year to year, and a smaller portion from acute episodes that lead to short-term spikes in resource use but are not followed by long-term needs.

A third determinant of the R^2-value is the *variation in explanatory variables*. A greater variation in the factors explaining variation in health spending (see Figure 3) ceteris paribus increases total variation. Whether the

proportion of predictable variation (R^2) then increases or decreases depends on whether the variation in explaining factors is known ex-ante or can be accurately predicted. For instance, greater variation in practice style, supply or input prices which are stable over time, enlarges both total variation in expenditures and the R^2-value. However, greater variation in input prices resulting from unpredictable changes in market power or government regulation, increases total variation but decreases R^2.

Fourth we hypothesize a positive relation between R^2 and the level of *medical technology*. This level may change from country to country and, within a country, it may change over time. An increase of the level of diagnostic technology may result in a better prediction of future (genetically determined) diseases and expenditures. In addition it may result in more protocolized treatments and thereby reduce random variation in treatments. An increase of the level of effective therapeutic medical technologies may keep alive at-risk patients who otherwise would have died, e.g., cancer patients, heart patients and patients with a transplantation. As a result the proportion of chronically ill persons may increase. As Kronick et al. (1995) stated, the expenses of the chronically ill are relatively more predictable because they pertain to chronic problems and their complications which persist from year to year.

A fifth determinant of the R^2-values that are presented in the literature is the *year of the data*. Many studies analyze data from the 1970s, while others use data from the 1990s. For the following reasons, which may be mutually related, we expect an increase in R^2 over time. First, medical technology has increased. Second, we have seen a substitution of outpatient care for inpatient care over the last decades, with outpatient care being more predictive than inpatient care. Third, the proportion of expenditures spent on prescribed drugs has increased over the last decades, with prescription drugs costs being relatively more predictable (see section 3.2.6). Fourth, the proportion of elderly and chronically ill persons, whose expenditures are more predictable, has increased. Fifth, the predictive power of age has increased over time. For example, Schut (1995) calculated that in the Netherlands from 1979 to 1986 the average hospital costs of men over 80 years old increased from 4.9 to 7.6 times the average hospital cost of men in the 45-49 years age group. So, over time more variation in expenditures can be explained by age. Based on these arguments we hypothesize an increase of R^2-values over time.

A sixth and final determinant of the R^2 values that are presented in the literature is the *length of the time period being predicted*. Using longer periods averages out some of the randomness, and tends to improve predictive power. Ellis and Ash (1989) developed models that predict a one month prediction period that achieved an R^2 of only .0089 on the monthly observations versus .04 using the same information with an annual prediction period. Alan Garber, Mark MaCurdy, and Mark McClellan (1998) examined the predictability of health spending over multiple years, and demonstrate the effects of smoothing out of random variation.

3.2.6. Maximum R^2

In an important set of articles Newhouse et al.[43] and Van Vliet (1992) ask the question: what is the maximum potential variance predictable by prospective risk adjustment models, i.e. models using only information from a past period or periods? The literature usually tries to answer this question by dividing the variance in actual spending into different components. The component indicating the between-person variance was estimated by McCall and Wai (1983) to be 0.15 and by Newhouse et al. (1989) at 0.145. Additionally, some within-person variance is predictable because of the autoregressive error-component [Newhouse (1996)]. As an upper bound for this component, exclusive of time-varying covariates, 0.04 could be used [44], making the predictable proportion around 0.20. This corresponds with the 0.174 estimated by Van Vliet (1992), who used an autoregressive moving averages (ARMA) model. However, the "around 20 per cent" is a lower bound on the ability to predict future spending because other predictive factors may be observed that are not reflected in past spending. (So it is a "lower bound on the upper bound", rather than a true upper bound.) Examples of such predictive factors are a pregnancy, a recent diagnosis of cancer, a terminal illness, or being on the waiting list for an expensive treatment [Newhouse et al. (1989, 1997); Van Vliet (1992)]. Plans and individuals could potentially predict more than the 20 percent of actual variance, but how much more is unclear.

[43] Newhouse (1996) and Newhouse et al. (1989, 1993, 1997).
[44] This value is based on Newhouse (1996). We consider 0.04 as an upper bound because Table 3 of Newhouse et al. (1989) contains correlations between expenditures and not between residual spending, as stated in footnote 62 of Newhouse (1996).

Results about the maximum R^2 as presented in the literature are consistent with the above mentioned determinants of R^2. Newhouse et al. (1989) estimated the maximum R^2 for inpatient care to be 0.08 and for outpatient care, 0.48. Similar results were found by Van Vliet (1992), who also concluded that the expenditures for prescription drugs together with GP consultations are extremely predictable (maximum R^2 of 0.80). This finding has serious implications for comparing R^2-values from a setting where expenditures for prescription drugs are not included (e.g., US Medicare data) with a setting where they are included (e.g., the Netherlands sickness fund data).

With respect to the relation between R^2 and subgroups Van Vliet (1992) found evidence supporting the hypothesis that predictability increases with age and that differences in health expenditures for older individuals are more predictable than those for young people. This hypothesis is consistent with the findings by Newhouse et al. (1989,1993) that the maximum R^2 for outpatient expenditures are higher for the age-group 14-64 years (maximum $R^2 = 0.48$) than for the age-group 3-13 years (maximum $R^2 = 0.37$).

Because of the relation between R^2 and both medical technology and the variation in factors explaining the variation in expenditures (such as input prices, supply, practice style and benefit plan features) it is important to note that these determinants may strongly vary from country to country. So one should be careful to apply in one setting the maximum R^2 estimated in another setting. Ideally, to have a benchmark researchers should estimate the maximum R^2 on the same (longitudinal) data base that is used for analyzing risk adjusters.

The relation between R^2 and year of data analyzed is relevant for the interpretation of the above mentioned lower bound of maximum R^2 (around 20 per cent). This estimate is based on different data sets from the 1970s and early 1980s. Based on the above arguments we are not surprised to see higher lower bounds to be estimated on data of more recent years. For example, Lamers (1999) analyzed acute care expenditures, including prescription drugs, for Dutch sickness fund members[45] for the years 1992-1996. Using the ARMA model [see Van Vliet (1992)] she found a lower bound of the maximum R^2 of 0.33.[46]

We started with the question: what proportion of variance in expenditures is potentially predictable by a health plan? We may conclude that the maximum is, in any event, much less than 100 per cent because many health expenditures cannot be foreseen by either the individual or the health plan [Newhouse et al. (1989)]. Furthermore a lower bound of the maximum percentage can be estimated, which depends on the type of care, the (sub)population, and the specific setting and year. However, we do not know how much more variation is predictable than indicated by this lower bound.

3.2.7. How successful can risk adjustment be?

With respect to the success of risk adjustment two types of concern can be discerned: (1) can risk adjustment be *sufficiently* successful?; and (2) can risk adjustment be *too* successful? We discuss both.

Newhouse et al. (1997) raise the question of how close to perfect the formula must be to make plans' incentive and ability to seek favorable risks a *de minimus* problem. We share their view that a workable formula need not achieve the ideal, but that it is unknown how far from perfection will be sufficient.[47] As stated in section 2.6 an adequate risk adjustment formula should be such that health plans expect the transaction costs of cream skimming (including the loss of good reputation) to exceed its profits. A second reason why the variation in the risk-adjusted premium subsidies will not equal the maximum potential variation in predicted expenditures, is that the sponsor ideally will only compensate for variation in S-factors and not in N-factors [Van Vliet (1992)]. A third reason why an adequate risk adjustment formula need not be perfect, is that cream skimming strategies based on one-year savings may have longer-run opportunity costs. Beck and Zweifel (1998) present an example in which 50 per cent of 'bad' risks turn out to be good risks in the long run, while 20 per cent of the 'good' risks become bad ones [because of regression

[45] Disabled persons and chronically ill are over-represented among the Dutch sickness fund members.
[46] Van Vliet and Lamers (1998), analyzing the same data base, found an R^2 of 0.19 for a model with the risk-adjusters 3-year DCGs and three years of prior expenses.
[47] Under the assumption of lognormally distributed expenditures Newhouse et al. (1989) provided evidence that as explained variation improves, incentives to select do not diminish proportionately. From this finding Newhouse (1996) concludes that the risk adjustment formula must be close to perfect to reduce greatly the incentives to select. However, as Van de Ven et al. (1994) put forward, after correcting for the overestimation of the nonlinearity in Newhouse et al. (1989), the relation between the square root of explained variance (which, just a profits, is a linear function of predicted expenditures, rather than a quadratic function) and profits appears to be linear.

towards the mean; see also Welch (1985)]. Fourth, the sponsor could periodically adjust and improve the formula, thereby lessening the attractiveness of some selection strategies in the long-term. Fifth, when the sponsor improves the formula, not only will a health plan's potential profits from selection decrease, but also the standard deviation of its profits will increase [up to a factor of three; Van de Ven et al. (1994)], thereby reducing the attractiveness of selection strategies. So, an imperfect formula may be sufficient to make selection unimportant. However, how much imperfection a sponsor can permit, is an unanswered empirical question.

A second concern about the potential success of risk adjustment is the question: can risk adjustment be *too* successful? A formula based on age and gender, each divided into two groups, clearly creates large incentives for selection, by pooling heterogeneous people in the same groups. Assume that, in order to reduce these incentives, a sponsor refines the subgroups and replaces the two age-groups by 40,000 birthday-groups and replaces the two region-groups by 10,000 zipcode-groups. Assume further that each of these 400 million subgroups contains at most one individual. Although this birthday-zipcode formula largely reduces the incentives for selection (except for those who change plans, for newborns and for those who are expected to die), it also reduces the health plans' incentives for efficiency because of the large extent of cost-based reimbursement with a one period delay. Most sponsors will reject the birthday-zipcode formula because of inappropriate incentives. The birthday-zipcode formula also lacks robustness in the sense of stability of the weights over time, and it suffers from overfitting in the estimation model.[48]

In the discussion of alternate risk adjustment models, we mention the conventional R^2 measures in a few instances in order to convey an initial picture of the explanatory power of different sets of information. We present other reasons why R^2 can be misleading and difficult to compare as well as alternative measures of predictive power useful for assessing different modeling frameworks in a subsequent section.

3.3. Alternative risk adjustment models

Considerable research has been conducted on alternative risk adjustment models in many countries, using a wide range of information. We discuss these models in groups defined by the kind of data used for prediction: demographics only, prior year expenditures, diagnoses, information derived from prescription drugs, self-reported health and functional health status measures, mortality, and other types of information.

3.3.1. Demographic models

The most basic type of information used to adjust payments to health plans (or providers) are age and sex. Figures 4A through 4C illustrate that there are pronounced differences in expenditures across individuals by age and sex, that these patterns differ according to the country and sub-population studied. Among privately insured enrollees in one large cross section of US firms from 1992-93 (Figure 4A), average health expenditures on men aged 60-64 are $4,100 versus only $350 for females aged 5-9, an 11-fold difference. The corresponding numbers for Medicaid (Figure 4B) are $4,160 and $340, a 12-fold difference. As shown in Figure 4C, the distribution of health spending by age and sex in the Netherlands for a large cross-section of people with the same insurance plan, there is also a more than ten fold difference in average costs between the highest and lowest expenditures.[49]

Age and sex are easy to document and use for risk adjustment, are fair, and generally accepted by all parties involved.[50] Because the information is independent of medical care, and not readily gamed, it appears attractive in terms of incentives. The most serious drawback of age and sex as risk adjusters is simply that they are weak predictors of individual expenditures.[51]

---

[48] The birthday-zipcode formula illustrates the need to make a distinction between the R^2 for explanation (in this case: $R^2 = 1.0$) and prediction (with Efron's R^2 being negative).

[49] Spending in the Netherlands was converted to US currency using the 1998 exchange rate of 2 Dutch Guilders per US dollar. Over the age range from 0 to 64, there is somewhat less variation in health spending in the Netherlands than in the United States.

[50] Separate calculation of health insurance premiums by sex is generally considered acceptable in the US and elsewhere, while separate calculation by race or religion is generally not considered acceptable. Interestingly, charging different insurance premiums by gender for automobile, life and liability insurance in the USA is generally NOT considered acceptable, even though gender-based differences in expected costs may be at least as large in these other insurance markets as in health care.

[51] Altman et al. (1998) highlight the important nonlinear relationship between age and health spending, as shown here in Figures 4a to 4c. They note that if a health plan has an above average age, then even without enrollment changes the average health spending of the health plan's

3.3.2. Prior year expenditures

Because expenditures in one year are correlated with expenditure the following year - the correlation coefficient for total health expenditures is on the order of .2 to .3 - a simple proposal has often been made to regress expenditures in year two on year one expenditures (together with other demographic variables) and use this model for calculating risk-adjusted payments. Newhouse et al. (1989), Van de Ven and Van Vliet (1992) and Ash et al. (1998) have all estimated such models and typically find that spending an extra dollar on health care in year one "predicts" spending of $0.20 to $0.30 in year two. The R^2 from a regression that includes age, sex and prior year expenditures, is generally estimated to be in the range of .06 to .10, with two recent estimates being .073 [Van Vliet and Van de Ven (1992)] and .098 [Ash et al. (1998)]. These measures are a substantial improvement over demographic only models, and comparable to the predictive power achieved by diagnosis-based models or models that use self-reported health status measures.

Although the accuracy of prior year expenditures is reasonable compared to many alternatives, this approach is inferior to others according to several of the above criteria. The feasibility of using such a model is often a concern. In some cases it can be seen as requiring the sponsor to "assume the can opener," since a major reason why a risk-adjusted capitation payment rather than cost-based payment is used is precisely because health expenditures are difficult to measure or monitor. In the USA in particular, a growing number of health plans do not collect individual level cost information that can be used for calculating payment for specific conditions. Instead, many plans have subcontracts with provider groups that do not even require that service and cost information be shared with the plan. The absence of cost or charge information undermines the feasibility of its use for payment in some settings, such as HMOs in the USA.[52] However in other settings, such as the Netherlands, the feasibility requirement is met, since prior year expenditures are routinely available in the administration of the sickness funds.

Although prior year expenditures or utilization appears to be the best single predictor of an individual's future health expenditures, some argue that using it as a risk adjuster creates inappropriate incentives. Firstly, some differences in prior use among individuals could reflect differences in physician discretionary practice patterns (an N-type factor). Premium subsidies based on prior utilization would pay health plans without regard to the appropriateness of the care McClure (1984)]. Secondly, the premium subsidies would be based on an average relationship between prior use and subsequent medical expenditures. The expected future costs, however, may differ widely for persons with high prior use associated with chronic medical conditions in contrast to those with one-off acute conditions. This might lead to inappropriate provider incentives or to new selection problems [Beebe et al. (1985)].

Some providers and researchers also challenge the fairness of using prior expenditures to calculate payments [e.g., Lubitz (1987) and Porell and Turner (1990)]. The usual argument is that payments based on prior year expenditures reward plans for spending more on individual patients, and punish "well managed" plans that conserve on spending. However, this argument misses the fact that plans are still only compensated for a proportion of their spending on health services. Ellis and McGuire (1993) and Newhouse (1996, 1998) have argued that this may be a desirable practice to soften the incentives of a fully prospective system (a thought that we develop further below). One last argument against using prior utilization/expenditures as a risk adjuster is that it does not provide higher subsidies to individuals with medical problems who have not sought care.

3.3.3. Diagnosis-based risk adjustment

The potential equity and inefficiency problem of inappropriate incentives related to prior utilization as a risk adjuster may be reduced by combining prior utilization with diagnostic information. Since the early 1980's a considerable amount of research has developed risk adjustment models that use diagnoses from insurance claims to calculate risk-adjusted payments. The three most widely known classification systems are the Ambulatory Care Group (ACG) system developed at Johns Hopkins by Jonathan Weiner and colleagues [Weiner et al. (1991, 1996)], the Diagnostic Cost Group (DCG) family of models developed at Boston University and Health

---

enrollees will increase faster than the average because of this nonlinearity. They define this concept as adverse retention, and demonstrate in one sample that as much as biased enrollment and disenrollment, adverse retention may explain why costs of health plans diverge.

[52] An alternative may be to use imputed spending based on a price or fee schedule applied to observed utilization.

Economics Research by Arlene Ash, Randall Ellis, Gregory Pope and colleagues [Ash et al. (1989, 1998); Ellis et al. (1996a, 1996b); Pope et al. (1998a, 1998b, 1999)], and the Disability Payment System (DPS) developed by Richard Kronick, and Anthony Dreyfus, [Kronick, et al. (1996)] primarily for U.S. Medicaid disabled enrollees.[53] Although the above authors have led the development of these classification systems, the models themselves have also been applied by numerous other researchers, notably Van Vliet and van de Ven (1993) and Lamers (1998a) in the Netherlands. Although each of these systems has its own unique features, they share several characteristics that are worth highlighting.

The starting point for all diagnosis-based risk adjustment models is the concept that certain diagnoses predict of health care expenditures. Each of the three major diagnosis-based models begins by identifying a subset of all diagnoses that predict current or subsequent year resource use. Although the three models differ in how they choose their subset of diagnoses, each attempts to identify codes that are assigned only for encounters involving a professionally trained clinician. In particular, diagnoses appearing on laboratory, diagnostic testing, and medical supplies claims are uniformly not used in classifying individuals for prediction, on the grounds that they are less reliable than those assigned by clinicians.

Since there are approximately 15,000 valid International Classification of Diseases (ICD9) codes, it is intractable to classify individuals at this level of detail because in most cases there will be too few people with a diagnosis to properly calibrate a model. Each of the models therefore begins by grouping ICD-9 codes into more aggregated groups based on clinical, cost, and incentive considerations. The most refined versions of the ACG, DCG, and DPS systems begin by classifying diagnoses into a large number of diagnostic-based groups, then use these diagnostic groups to classify individuals according to the specific combination of conditions each individual has. The approaches that each model uses, and the way that information is used to generate predictions differ in the three models.

As described in Weiner et al. (1996), the Ambulatory Care Group methodology begins by classifying a subset of all valid ICD9-CM diagnostic codes into 32 diagnostic groups.[54] Depending on the model specified, various combinations of these diagnostic groups are then used to classify each individual into one of up to 83 mutually exclusive Ambulatory Care Groups. In a few cases the ACGs correspond to specific medical conditions (e.g., Asthma); however in most cases the groups are relatively broad ("Acute: Major", "Chronic Medical, Unstable" , "Chronic Specialty"). Thirteen of the groups are based on counts of how many of the 32 detailed diagnostic groups the patient has, and hence explicitly reward plans for coding more conditions. Payment weights, based on regression analysis, can be used together with ACG assignments to predict individual or group level resource use.

As suggested by their name, ACGs were originally designed to use only ambulatory diagnoses, and hence the ACG algorithms ignore inpatient episodes. Using different classification systems designed to incorporate inpatient diagnoses, the ACG framework has also been expanded to include inpatient diagnoses in two different ways. The first approach summarizes inpatient conditions using simply the 15 Major Diagnostic Categories (MDCs) into which Medicare program's Diagnosis Related Groups (DRGs) can be grouped. These MDC categories are nonspecific to severity differences within a broad body system ("Infectious and Parasitic Diseases", "Cancers", "Diseases of the Circulatory System", etc.). The second approach uses what Weiner et al. call "Hospital Dominant" (HOSPDOM) inpatient diagnoses. The model only recognizes inpatient diagnoses for which at least 50 percent of the patients with that condition were hospitalized in a benchmark dataset. The goal is to avoid rewarding plans for unnecessarily hospitalizing patients in order to increase payments. However this is a very strict criterion for deciding which diagnoses inpatient diagnoses will affect model predictions.

Although they are not based on the same classification system as the ACG system, the Payment Amounts for Capitated Systems (PACS) developed by Gerry Anderson et al. (1990) is an inpatient diagnosis based system developed using Medicare data. The most distinctive feature of the PACS system is that it counts how many times a person is hospitalized over a two year period within each of fifteen MDCs. It also notes whether the person has any outpatient visits in the year prior to the year being predicted, and classifies hospitalizations into four chronicity levels. The model does not address incentives or other economic, rather than clinical, criteria.

[53] Diagnosis-based risk adjustment models have also been developed by Hornbrook et al. (1991), Clark et al. (1995), and Carter et al. (1997).
[54] ACGs have recently been renamed "Adjusted Clinical Groups". See http://www.hsr.jhsph.edu/acg/acg.html for references and further details about ACGs.

The Diagnostic Cost Group (DCG) risk adjustment models were originally developed by Arlene Ash et al. (1989) using data from the US Medicare population from 1979-80. At the time of its early development, diagnostic information was not yet routinely coded on outpatient claims, and there were also concerns about the completeness of secondary diagnostic codes even on inpatient records. Therefore, the earliest versions of the DCG models used only principal inpatient diagnoses. A series of reports and papers summarized in Ellis and Ash (1995) built upon this early work, used data from the mid 1980's, and explored a variety of extensions that continued to use only principal inpatient diagnoses.

Early DCG models are "single hierarchy" models. Modeling begins by clustering diagnoses into a large number of clinically homogenous groups. These diagnostic groups were then further aggregated into a small number (between 9 and 20) of Diagnostic Cost Groups (DCGs), according to empirically determined similarities in the future cost of individuals hospitalized with different diagnoses. Some diagnostic groups are ignored in the classification process because they are viewed as being too discretionary or too ambiguously coded. Individuals with multiple hospitalizations in a given year are uniquely assigned to the most expensive DCG in which any of their hospitalizations fell, thus establishing a "single hierarchy." Individual DCG scores are included as categorical variables in linear regression models and used to predict future costs.

A more recent series of studies by the same group [Ellis et al. (1996a, 1996b); Ash et al. (1998); Pope et al. (1998a, 1998b)] has significantly expanded the original DCG framework. One fundamental change is that instead of using only principal inpatient diagnoses, these recent DCG models use all diagnoses from encounters with clinically trained medical professionals, including secondary hospital inpatient diagnoses, hospital outpatient facility diagnoses, and other diagnoses assigned by clinicians. A second fundamental change is that instead of only noting the most serious diagnosis, the models capture multiple conditions. Instead of a single hierarchy used to rank all diagnostic groups, the recent models use information about multiple conditions, and impose hierarchies on diagnostic groups only when they are clinically related to each other. Numerous other important changes were also made. Considerably more clinical input was used to identify selected subsets of diagnoses to use in the first stage of the classification system. The system now includes 543 detailed diagnostic groups, which are further collapsed into 118 groups that are now called HCCs - Hierarchical Condition Categories. The populations studied were expanded from being simply Medicare enrollees to include privately insured and Medicaid eligibles. Instead of clustering diagnoses into cost groups before running a regression, selected diagnostic clusters - the HCCs- are included directly in regressions, so that estimated regression coefficients reflect the incremental cost of specific medical conditions. Concerns about discretionary admission and creating inappropriate incentives were incorporated by excluding selected HCCs from inclusion in regression models that predict subsequent year costs.[55]

In addition to the developmental work in Boston on DCGs, considerable exploration and further developments using the DCG framework have taken place in the Netherlands. Van Vliet and Van de Ven (1993) evaluated ten different risk adjustment models, including the original principal inpatient DCG models, DCG models that exclude certain diagnoses due to concerns about discretion noted above, and DCG models customized for the Netherlands. They also draw useful comparisons to models that are based on the PACS system using dummy variables for the MDCs in which each hospitalization falls, and models using prior year expenditures.

Lamers and Van Vliet (1996) expanded the DCG framework by considering multiple years of hospitalizations. The rationale for this is twofold. First, having had a serious hospitalization in a given year might induce predictably above-average expenditures not only in the year directly following but also, to a diminishing degree, in the years thereafter (without necessarily resulting from a new hospitalization). Secondly, by giving higher premium subsidies for people who have been hospitalized for certain diagnoses during one of the previous years (instead of only during the last year), the probability increases that a health plan will receive an appropriate premium for its chronically ill enrollees. The results of Lamers and Van Vliet indicate that the predictive accuracy improves when DCGs over a longer period are incorporated in the subsidy formula. For example, for the five per cent enrollees with the highest costs in year t-4 the predictable losses in year t decreased from 88 per cent (demographic model) to 62 per cent (1-year DCG model) and to 43 per cent (3-year DCG model) of the predicted costs.

[55] See http://www.DxCG.com for references and further discussion about DCG models.

Although the DCG-models outperform a model based on age and gender only, there still exist subgroups with substantial predictable losses. Lamers (1998b) showed that when the sponsor uses a (1-year or 3-year) DCG-model, a group of about 30 per cent "bad risks" can be formed by using selected information from a health survey, such as perceived health, having functional disabilities, consultation of the general practitioner, use of home nursing and the number of prescribed drugs. These "bad risks" on average have a predictable loss of more than half the overall mean per capita expenditures. More than 90 per cent of the "bad risks" were not hospitalized in the previous year.

The Disability Payment System (DPS) of Kronick and Dreyfus was developed with the specific aim of risk adjusting payments for persons eligible for Medicaid by reason of medical disability. The DPS system is similar to the DCG/HCC system in using all diagnoses from clinical encounters, incorporating hierarchies and concern about incentives, and explaining particularly well the upper tail of the health expenditure distribution. It is somewhat more additive than the DCG/HCC model, taking note of how many conditions a person may have within certain body systems. Because the DPS system has mainly been used for persons with disabilities, it is not clear how well it works for other population subgroups.

One important advantage of diagnosis based risk adjustment is that data often exist for large samples on which models can be developed and calibrated. Although diagnosis based risk adjusters tend to do well in predictive accuracy and feasibility, they do less well on fairness to providers or plans with different levels of completeness in recording diagnoses. Diagnosis based systems almost invariably reward plans that more actively encourage patients to seek treatment. For instance, if a plan screens more aggressively for certain conditions, then they are more likely to detect them, hence increasing payments. Similarly, a plan coding baby deliveries so as to justify performing more Caesarian deliveries will tend to have an enrollee mix that looks sicker than a plan that does not encourage Caesarians. The distortionary effects of using diagnoses for risk adjustment is potentially compounded if the risk adjustment system only notes hospitalizations: if a patient will only be eligible for a higher payment if hospitalized for a condition predictive of higher subsequent year costs (e.g., HIV/AIDS or colon cancer) then a risk adjustment system based on only hospital diagnoses (and not on outpatient care) will encourage unnecessary hospital admissions. This undesired incentive is smaller in a 3-year DCG-model than in a 1-year DCG-model.

3.3.4. Information derived from prescription drugs

Another approach for extracting health status information from prior utilization data is to infer the presence of chronic conditions from the use of prescription drugs. Since pharmacy information is often available with a short time lag, this is another attraction of drug information. Hornbrook et al. (1991) classified drugs into different therapeutic classes. In each class the number of drug orders was counted. Adding 19 drug classes to the adjusters age and gender yielded an increase in R^2 from 0.021 to 0.050. Von Korff et al. (1992) used outpatient pharmacy data to develop the Chronic Disease Score (CDS). The CDS weights were based on physician judgment of disease severity. This CDS was found to predict hospitalization and mortality after controlling for age, gender and health care visits. Clark et al. (1995) revised the CDS by empirically estimating the weights for individual drug classes. They distinguished 28 different conditions. By adding 28 CDS dummy variables as additional risk adjusters to age and gender the predicted variation in total medical expenditures of adults in the next half year increased from 3 per cent to 10 per cent. The results of Clark et al. also suggest that adding information derived from ambulatory diagnoses to the revised CDS adds little additional explanatory power.

Lamers (1998c) built on the revised CDS developed by Clark et al. To prevent manipulation she put "alike" conditions (for example hypertension and cardiac disease) in the same group. The chronic conditions could be clustered into six so-called *Pharmacy Cost Groups* (PCGs) on the basis of empirically determined similarities in future costs without affecting the predictive accuracy of the model. Lamers concludes that although PCGs are good predictors of future health care costs, their usefulness as risk adjusters may be restricted because of inappropriate incentives. The additional subsidy for a PCG-classified enrollee (far) exceeds the costs of the prescribed drugs that form the basis for PCG-assignment[56]. A similar result was reported by Ellis (1985), who found that each dollar spent on drugs predicts $3.73 of health care expenditures the following year. Given the high predictive value of CDS and

[56] Given the conclusion by Clark et al. (1995) about the similarity in predictive value of the revised CDS and information derived from ambulatory diagnoses, it is interesting to know whether such perverse incentives also exist in case of risk adjusters based on outpatient diagnostic information.

PCGs future research should be directed at minimizing the information surplus of health plans that have access to information about prescription drugs of their enrollees.

3.3.5. Self-reported health information

A fundamentally different approach to risk adjustment is using self-reported measures derived from surveys. Survey-based information has several advantages over diagnosis-based systems [see e.g., Gruenberg et al. (1996); Hornbrook and Goodman (1996)]: most information is not contingent on having come in contact with a medical provider, no prior history of claims or enrollment is needed to generate predictions, measurement of consumer perceptions of need and anticipated use, uniformity across health plans, and socioeconomic (lifestyle, taste, employment) variables can be measured in addition to health status. There are also some important disadvantages of self-reported measures. Surveys are relatively costly to collect (typical numbers for the USA range upwards from $30 per completed survey). Response rates can be unacceptably low and correlated with medical risk. Large samples generally do not exist on which to develop reliable prediction models. Some survey questions raise confidentiality and accuracy concerns (e.g., questions about HIV/AIDS or mental illness). Although surveys do not require providers or health plans to provide claims information, in many cases health plans are expected to assist with implementation of the surveys, raising concerns about nonrandom sampling or follow up. As shown below, self reported measures generally do not have as high an explanatory power as diagnosis based systems.

The most common type of information collected through surveys is perceived health status. In its simplest form, it can be a single self-reported health summary of excellent/very good/good/fair/poor. Asking how health status has changed since a year earlier also can be significant. More elaborate surveys such as the Short Form 36 (SF36) [Thomas and Lichtenstein (1986); Ware and Sherbourne (1992)] or the closely related RAND-36 survey [Hornbrook and Goodman (1995)], measure perceived health status along eight dimensions. A second class of information is functional health status, which assesses how well can the individual perform various activities. The two most common instruments are the Activities of Daily Living (ADLs) and Instrumental Activities of Daily Living (IADLs). A third class of self-reported measures relates to chronic conditions (e.g., diabetes, high blood pressure, asthma, etc.). Such measures, while not coded by a physician, may require that the individual has received a diagnosis from a clinician, and hence to that extent depend on contact with providers. Other self-reported measures include information such as lifestyle (smoking, drinking, food), marital status, employment, education, and whether a person can drive.

Several studies have compared the predictive power of self-reported and diagnosis based risk adjustment models. Table 3 presents R^2 measures from six studies that included various self reported measures. While all of the models that incorporate self reported measures are superior to models that include only age and sex, the self reported models have lower R^2 than the models that include diagnostic information. We note that all of the sample sizes are relatively small, so that overfitting is a concern. Only Lamers and Pope et al. present R^2 which are based on out-of-sample predictions, and hence are robust to overfitting.

Table 3

3.3.6. Mortality

Mortality has been suggested as an additional risk adjuster because of the high health care expenditures prior to death [see e.g., Tolley and Manton (1984); Lubitz (1987)]. There are different opinions about its usefulness. Van Vliet and Lamers (1998) conclude that mortality should not be used as a risk adjuster. Their argument is that most of the excess costs associated with the high costs of dying are unpredictable. Even with their most comprehensive regression model (R^2= 0.189; acute care, general population) the actual costs of decedents are still 250 per cent above predicted costs. Furthermore, they found that in the Dutch situation the allocative effect of a mortality-based prospective adjustment based on standardized mortality ratios for the past 5 years, would be very modest: a sickness fund with an extreme excess mortality of 10 per cent could expect an increase of its age-gender based premium subsidies of about 0.25 per cent. One may also wonder whether it is politically and socially acceptable for a health plan to receive a higher subsidy when more of its members die.

In the US, Tolley and Manton (1984) studied the possibility of using cost-weighed cause-specific local mortality for setting Medicare-subsidies to at-risk HMO's. They concluded that it would require a difference in mortality

experience of more than 30 per cent to result in a 6 per cent change in the mortality-adjusted AAPCC. Van Vliet and Lamers (1998) note that, although cause-of-death information is theoretically attractive, practical concerns include reliability, validity, availability, manipulation, auditing and privacy of the data.

Another opinion about mortality as an additional risk adjuster is expressed by Beck and Zweifel (1998). They conclude that a dummy variable indicating death during the observation period should be included in the subsidy formula. They suggest to retrospectively compensate health plans with a prospectively determined payment per death.

The only country, as far as we know, that applies mortality as a risk adjuster is Belgium. The mortality adjuster in Belgium is based on the average number of deaths per 1000 enrollees in prior years at the health plan level [Schokkaert et al. (1996)].

3.3.7. Models using other information

A wide array of alternative risk adjustment information has been examined in the literature. See Epstein and Cumella (1988) for an early review. Van Vliet and van de Ven (1992) evaluated many demographic variables including employment, family size, region and found that these additional explanatory variables improved the R^2 of .028 from an age sex model to only .037 (see table 3). It is interesting to note that regional dummies increased the R^2 by .006 from .028 to .034. This small improvement may reflect that Netherlands has less regional variation than many other countries.

Disability and functional health status have been shown to be relatively good predictors of future expenditures [Thomas and Lichtenstein (1986); Hornbrook and Goodman (1996)], Indicators of functional health status reflect someone's ability to perform various activities of daily living and the degree of infirmity. Disabled and functionally impaired persons appeared to have roughly twice the health care expenditures of those who are unimpaired [Lubitz et al. (1985); Gruenberg et al. (1989)]. Impairment level continued to be a significant contributor to high Medicare expenditures after controlling for demographic factors and prior utilization [Gruenberg et al. (1989)]. Newhouse (1986) considered disability to be an almost ideal adjuster. In 1998 an indicator of disability is used as a risk adjuster in Belgium, Germany and the Netherlands.

As highlighted in section 2.1.2, input price variation is a cost factor that is likely to be attractive to include in risk adjustment formulas. In the United States, counties are used as the geographic basis for risk adjustment under the 1998 Medicare HMO payment formulas. Instead of using a county based price index, however, the Medicare program in the US in 1998 used a function of the county level average indemnity payments, by age - sex groups to calculate payments. These county averages reflect not only input price variation, but also practice style variation. Hence they almost certainly overstate the variation that policy makers should use in calculating risk-adjusted subsidies.

Figure 5, based on data in Ellis et al. (1996b), highlights that much of the geographic variation in the US Medicare capitation payment is systematically related to differences in input price variation, although clearly other factors are also varying geographically. The scatter plot highlights that there is a considerable amount of the variation in average costs across metropolitan areas that is explainable by factor price variation. (The simple correlation coefficient between input prices and average costs is .61 ($p < .001$)). Based on this evidence there appears to be a clear rationale for geographic adjustment for input price differences.

3.4. Predictive power

3.4.1. Issues with R-square measures

This chapter has so far presented only the conventional R^2 when contrasting the predictive power of different models. In this section, while comparing selected models more fully, we highlight three factors that influence the R^2 that have not always been emphasized in the risk adjustment literature: the impact of truncating the dependent variable, the impact of a logarithmic transformation, and the possibility of overfitting models in small sample sizes. We then present selected additional ways of assessing risk adjustment models.

Figure 6 presents results from five different studies that used different truncation points of the dependent variable. Truncation is also sometimes called top-coding, to distinguish it from censoring, which is the dropping of observations, and is also sometimes done. Whether truncation is appropriate depends on how the risk adjustment system will handle "high outliers." Truncation is often justified on the basis that health plans may reinsure, in which case very high costs really are not borne by the health plan. While this may be appropriate, it can muddle comparisons across different studies.

Figure 6 makes the important point that truncating health expenditures at a maximum level can have a major impact on predictive power. The predictive power of all of the models increase on the order of 30-70 percent once the upper tail of the expenditure distribution is truncated. Note that the difference of truncation is less pronounced with the self reported measures, and appears to be the greatest for the diagnosis based models. The comparisons made between DCGs and ACGs by Dunn et al. (1996) suggests that DCGs and ACGs do similarly if health spending is truncated at $25,000, but that DCG models achieve a higher R^2 if one uses untruncated spending.

A second factor complicating comparisons across different studies is that many studies use the natural logarithm of expenditures for some or all of their calculations of R^2. Figure 7 highlights that using a log transformation inflates the conventional R^2 by about 100 percent, and this holds whether one is using age sex models, self-reported or diagnostic information. As highlighted above, Manning (1998) has demonstrated that the log transformation has undesirable statistical properties for predicting spending in absolute levels, and is difficult to use for generating unbiased predictions even with retransformation.

Figure 8 highlights a problem that has only recently been presented systematically. Ellis and Azzone (1998) used Monte Carlo draws of various sizes from a larger sample to demonstrate that the ordinary R^2 is systematically overstated, even when relatively large samples of 10-50,000 people are used. The problem is more pronounced for the more detailed models such as the DCG/HCC model using all diagnostic categories. Yet even simple age-sex models overfit the data for small sample sizes, and overstate the R^2. (Note that the R^2 for the age-sex models has been multiplied by a factor of ten to facilitate being able to see the result on the same scale). The implication of this finding is that R^2 from diagnosis based models estimated on samples of 50,000 without a (repeated) split sample analysis should be deflated by a factor of 1.3 when comparing to models based on a million observations.

Instead of simply using the individual R^2, Ash et al. (1989) developed the concept of a Grouped R^2. Instead of squaring the difference between actual and predicted spending for each individual, they explored methods of summarizing predictive power using the weighted squared deviations for the averages of exhaustive subgroups. Using this technique, Ash and Byrne-Logan (1998) found that diagnostic based models such as ACGs and DCGs can explain about 80 percent of the variation that is explainable by prior year spending once prior year spending is summarized by averaging people into 50 equal-sized samples of 384 people after sorting by year one spending. This measure results in much higher estimates of predictive power because it averages out much of the random error.

3.4.2. Comparisons using other than R^2

The conventional R^2 attaches enormous weight to large outliers: the one person in a sample costing a million dollars more than expected will add as much to the variance as 1,000,000 people with prediction errors of $1,000. To offset this, several researchers have also presented comparisons of different models using the Mean Absolute Error (MAE) [Dunn et al. (1996); Ellis et al. (1996b); Ettner et al. (1998)]. The main difficulty in using this measure is that it is less commonly used, and hence it is more difficult to assess what a good level or improvement of the MAE is.[57]

A further approach that is widely used is to assess the predictive power of various models using selected subsamples of the population being predicted. For example, actual expenditures can be compared to predicted expenditures using each model, for either random or nonrandom subsamples. Comparisons using randomly sampled groups of people provide information about stability of payments and their standard errors. Given that even the best risk adjustment models leaves perhaps 90 percent of the variation unexplained, once samples reach sizes such as 5,000 enrollees, differences between the predictive power of risk-adjusted versus non-risk-adjusted

[57] Also, from the perspective of assuring access for sicker people, larger errors may be disproportionately important.

payments *for random samples* are hardly worth noting. Ingber (1998) provides a good example of the use of this approach.

Given that one of the most important criteria to use in selecting among risk adjustment models is the incentives they create for selecting or dumping certain types of people, one of the most useful assessments that can be done is to compare actual and predicted expenditures for selected nonrandom groups of interest. Figures 9, 10, and 11 present three such comparisons using a privately insured sample from the US from Ash et al. (1998). Figure 9 provides an out-of-sample comparison of actual and predicted expenditures for ten chronic diseases, as identified from insurance claims from inpatient and outpatient bills from hospitals, physicians and other clinically trained professionals. As is readily seen, an age-sex model does not distinguish among the high cost chronic conditions, and will not reduce incentives to avoid enrolling such individuals. In contrast, a diagnosis based risk adjustment model, such as the DCG/HCC model shown here, can pick up a substantial amount of the variation across different chronic disease groups.[58]

Figure 10 compares predicted and actual spending for nonrandom groups, defined by sorting the sample in ascending order by the predicted level of spending in year two. In Figure 10, this comparison is made using predicted costs from the DCG/HCC model in a privately insured sample. Such a figure is useful for assessing whether the risk adjustment model can identify both very high cost and very low cost individuals, and whether the people predicted to be high or low cost do in fact incur these costs. The risk model predicts people with expected costs that range from $270 in the lowest cost interval, to $51,962 in the highest cost group, a factor of nearly 200 to 1. The DCG/HCC predictions track differences in average actual spending quite well for groups defined in this way. In contrast the age-sex model (whose average predicted costs are also plotted) varies by a factor less than ten to one.

While Figure 10 is informative about how well a model can identify high and low cost people, it is likely to be biased in favor of the risk adjustment model that is used to create the intervals shown on the horizontal axis. Since the prediction groups shown on the horizontal axis are defined using groupings from the DCG/HCC model, they are likely to make the DCG model look better than any other model whose predictions are plotted against the DCG defined groups, such as the age-sex model. In a similar way, groups that are defined using self-reported measures are biased in favor of making self reported measures look good, and groups defined over prior year spending are biased in favor of making prior year spending models look good. [See Pope et al. (1999), and Ash and Byrne Logan (1998) for a discussion of this issue.]

Figure 11 presents another way of examining two competing models on neutral territory. Here the population is divided into 50 same-sized groups, based on actual year 1 spending. Each observation is defined by sorting people by year 1 actual spending, and then dividing them into 50 groups. For each group, average year 2 spending is plotted against average predicted spending. The figure compares the predictions of Ambulatory Care Groups (ACGs) with Diagnostic Cost Groups (DCGs). The population is an independent sample of 192,000 privately insured people covered by the Massachusetts Group Insurance Commission [Ash and Byrne-Logan (1998)]. Here ACGs do better than DCGs for many of the low cost groups, but worse for the highest two percent.

3.5. Directions of ongoing development

Thus far the risk adjustment models discussed have used information to predict individual annual health spending for broad population groups. Several other directions of empirical research have been explored in the literature. We group them into three approaches: those that use information to predict less than annual patterns of spending, those that predict only specific subpopulations, and those that carve out certain services. Finally we discuss ongoing developments around conventional versus optimal risk adjustment.

3.5.1. Timing information

There is potentially a considerable amount of information contained in the date at which new diagnostic or other

[58] The chronic disease groups shown in this figure were selected by HCFA in 1995, and overlap with, but do not coincide with the classification system used by the DCG system. The same chronic groups have been used in Ellis et al. (1996a), Weiner et al. (1996), Ash et al. (1998), and Pope et al. (1998a, 1998b, 1999).

claims based information is coded. This information is not utilized by any of the risk adjustment models currently in use. For instance, both the ACG and DCG models use diagnostic information without distinguishing whether a person has the diagnosis at the beginning or the end of the base period. Information arriving near the end of the base period will in general be more predictive of spending patterns the following year, and hence it is easy to imagine using this information in making predictions. Ellis and Ash (1989) developed a "Continuous Update DCG model" which used information from a twelve month base period to predict ahead only one month. By making a series of twelve one month predictions, they were able to substantially improve upon the overall predictive power of a prospective model. When the twelve monthly predictions are aggregated to a calendar year, then the resulting DCG models, using only the principal inpatient diagnoses, achieved an R^2 that was more than double that of the simpler model that predicted a full year ahead (.089 versus .039). Ellis et al. (1996a) extended this to use all diagnoses and achieved an R^2 of .24. Using a two step estimation algorithm, Ellis (1990) developed a "Time Dependent DCG model" using principal inpatient diagnoses in which nonlinear time profiles for each diagnostic group were estimated. This model also achieved an R^2 of .24.

3.5.2. Selected subpopulations

A different dimension of research has been to develop risk adjusters for selected subpopulations. For example, Weiner et al. (1991, 1996) have estimated their ACG models for privately insured, Medicare, and Medicaid populations in the US. Ash et al. (1998) and Pope et al. (1998a) have also done so for DCGs. As previously discussed, the Disability Payment System (DPS) of Kronick et al. (1996) has been developed primarily for Medicaid populations eligible for reasons of disability. Risk adjustment models have also been estimated separately for pediatric populations [Newhouse et al. (1993), Fowler and Anderson, (1996)], for persons with HIV/AIDS [Kahn et al. (1995)], and for end stage renal disease (ESRD) patients [Farley et al. (1996)]. The rationale for developing of separate models for each of these groups is that they are vulnerable populations. Incentives under a system that does not specifically distinguish among the characteristics of these subgroups may result in health plans not wanting to enroll them. To the extent that they are unprofitable, or their costs are more uncertain, then there is also a greater risk that payments will be unfair.

3.5.3. Carve outs

Another dimension for ongoing research has been to predict costs for specific sets of medical services. Services such as pharmacy costs, behavioral health care (mental health and substance abuse treatment), dental coverage, and neonatal costs are frequently "carved out" of the expenditures being subsidized. The economic rationale for carve outs can be related to both demand side considerations (they may be harder to manage under a capitated system) or supply side incentive concerns (the carved out services may be more vulnerable to underprovision as health plans attempt to attract profitable enrollees). Frank et al. (1997) and Ettner et al. (1998) develop the rationale for carveouts and separate risk adjustment formulas for behavioral health services in light of the danger that selection incentives will be particularly strong in this group.

3.5.4. Conventional versus optimal risk adjustment

An important dimension of ongoing research is based on the notion that there are two broad approaches to calculating risk adjusters: statistical and economic. More recently, this has been given the names "conventional" versus "optimal" risk adjustment [Glazer and McGuire (forthcoming)]. There is a small but rapidly growing literature that is examining optimal risk adjustment and over the next decade may dominate risk adjustment research.

Conventional risk adjustment modeling has focused primarily on how well various risk factors can predict current health spending. In most cases regression models are estimated so as to generate unbiased predictions conditional on available information. If risk selection is viewed as a problem, as it has been in the U.S. Medicare program, then the sponsor may choose to pay some proportion of the predicted amounts, so as to recover some of the distortion or cost savings from selection efforts. But a key implicit assumption of conventional risk adjustment is that the pattern and level of health spending on a given individual is exogenous, and is not itself affected by the risk adjustment formula.

Another way to think about risk adjusters is to regard them as prices that can be set by regulation, in order to achieve efficiency and equity objectives. In a new literature, researchers are making explicit assumptions about the objective of risk adjustment policy, and the market conditions in which risk adjustment takes place, in order to characterize “optimal’

risk adjustment. This optimal risk adjustment literature asks questions such as, "If X is the set of the risk adjusters feasible to use, what are the optimal weights on these adjusters to minimize selection-related inefficiencies." In general, the answer need not be the conditional means that come from statistical research. In an analysis of adverse selection, Glazer and McGuire develop a model in which there is a supply-side moral hazard problem: plans can distort the services offered in order to attract profitable enrollees. They show that if an insurance market contains any element of "separation" of risks in equilibrium, risk adjustment can be improved over statistical average risk adjustment by putting more weight (paying more) for adjusters associated with high costs. In their example, age, an imperfect signal of true severity, is available for risk adjustment. Conventional risk adjustment on age overpays for the healthy and underpays for the sick (generating the selection problem). The plan attracting the sick will be providing too few services in equilibrium. The regulator can do something about this because the sick persons' plan has more old people. By paying more for the elderly, the regulator can increase the spending on the sick. (Corresponding efficiency gains appear in the plan for the healthy too, who may no longer need to separate themselves from the sick.) The new insight of their work is that socially optimal risk-adjusted payments should not simply reflect expected costs, but should also reflect the demand side process through which patients choose health plans and providers, as well as the supply side process through which health plans adjust the service mix to attract specific types of patients.

Notwithstanding the theoretical elegance of their work, its practical relevance depends on the strength of their assumptions. Glazer and McGuire assume that health plans select the profitable enrollees *only* via the distorting services offered and not via other tools for selection (such as those mentioned in section 2.4.2). If health plans use these other tools, which is quite realistic, paying more for the elderly and less for the young may create a new selection problem: health plans will strive to avoid the young. It also makes it even more rewarding for a health plan to select the healthy elderly (e.g., by selective advertising using the addresses of the fitness club for elderly) than it already was. Further, their assumption of a positive correlation between age and risk factors that are known to the consumers but for which the subsidies are not adjusted, may not always be fulfilled in practice. It may be true for diabetes and cancer, but not for AIDS (young men) and neonatal care (which parents may anticipate because of genetic information or tests). So distorting payments based on an imperfect signal could potentially increase the selection against certain types of patients.

Glazer and McGuire (forthcoming) develop a Bayesian framework in which the sponsor uses an imperfect signal to detect among two patient types, and distorts along a single dimension of services. Frank, Glazer and McGuire (1998) extend this work by examining its empirical implications with multiple services. They show that the sponsor should optimally take into account not only the marginal expected cost of each service, but also the predictability of the service, and how highly correlated the service is with other services. Their empirical specification suggests that the services most vulnerable to being under-supplied are those which predict high future costs, have less uncertainty about this prediction, and are highly correlated with total spending. They use this framework to identify services such as mental health and pharmaceuticals that are most prone to undersupply, but do not try to empirically estimate whether in fact these services are particularly undersupplied. They also show that the method used to risk adjust health plan payments can influence the incentives plans have to distort services.

Shen and Ellis (1998) develop a model in which plans are able to act on private information that cannot be used for risk adjustment by the sponsor. (For example, a plan may know that an individual is hypochondriac.) In their model, plans can perfectly cream skim, but there is no moral hazard problem, i.e., individual expenditures are exogenous characteristics of consumers. They show that conventional risk adjustment can be improved upon in this scenario as well, although interestingly in the opposite direction from Glazer and McGuire. They show that in order to minimize total health payments by the sponsor, payments to plans attracting a favorable selection should be considerably less than that implied by conventional risk adjustment using imperfect risk adjusters.

In one more recent paper from the optimal risk adjustment literature, Encinosa (1998) examines optimal risk adjustment in a world with both moral hazard (HMOs can choose effort) and cream-skimming (HMOs can identify and perfectly select low risk types in the population). He demonstrates not only that risk adjustment may not be able to achieve the social optimum, but that under certain conditions conventional risk adjustment can be worse than no risk adjustment if there is market power by health plans (he examines the duopoly case). It is disturbing, but not surprising that risk adjustment cannot achieve the first best in the presence of market power.

The hallmark of this optimal risk adjustment literature is that conventional risk adjustment can lead to biased predictions of actual spending. This bias can arise either because health plans distort spending patterns, so that spending is not

exogenous to payments, or because plans have private information or can otherwise distort enrollment, so that the observed risk factors are biased predictors of actual costs. While a great deal of research remains to be done in this area, we speculate that there may be a reconciliation between conventional and optimal risk adjustment which is that once the behavioral response by the plans has taken place, in general it will be necessary to recalibrate the risk adjustment model to reflect actual spending patterns *within the competing, risk-adjusted health plans*, rather than spending patterns from before risk adjustment occurred, or from some fee-for-service sector.

4. Risk sharing

If the risk-adjusted premium subsidies are not sufficiently refined to reduce selection, a complementary strategy is risk sharing between the sponsor and the health plans [Gruenberg et al. (1986); Newhouse, (1986)]. *Risk sharing* implies that the health plans are retrospectively reimbursed by the sponsor for some of the acceptable costs[59] of some of their members. Consequently the risk-adjusted premium subsidies have to be adjusted to the health plans' new financial risk. There is a clear analogy between such risk sharing and the outlier payments in the system of diagnosis-related group (DRG-)payments to hospitals [see e.g., Keeler et al. (1988)]. Although risk sharing effectively reduces a health plan's incentive for selection, it also reduces its incentive for efficiency. So, given some restrictions related to the premium contribution and given an open enrollment requirement, there is a tradeoff between selection and efficiency [Newhouse (1996)]. The goal of risk sharing as discussed in this section is to reduce the health plans' *predictable* losses and profits, while preserving their incentives for efficiency as much as possible. It is not the goal of risk sharing to reduce a health plan's financial risk by reducing the random variation of its expenditures. This may be achieved by traditional reinsurance.

An essential difference between traditional reinsurance and risk sharing as discussed in this chapter, is that for reinsurance a health plan has to pay a *risk-adjusted* premium to the reinsurer. Consequently, reinsurance does not reduce the health plan's predictable losses on high-risk individuals. It even increases them because of the loading fee included in the reinsurance premium. Therefore, traditional reinsurance cannot be a tool to reduce the health plans' incentives for selection[60]. Risk sharing, as discussed here, could be described as a "mandatory reinsurance program with regulated reinsurance premiums" as distinct from voluntary reinsurance with risk-adjusted reinsurance premiums. The retrospective payments from the sponsor are comparable with a reinsurer's retrospective payments, and the payment that a health plan forgoes because of the financing of the sponsor's retrospective payments (see next paragraph) can be considered a "mandatory reinsurance premium".

There are at least three ways to finance the sponsor's retrospective payments to the health plans: First, the sponsor may reduce the premium subsidy. The reduction of the subsidy per risk-group could be equal to the mean per capita predicted ex post payments that all health plans together receive for consumers in this risk-group. Alternatively, all premium subsidies could be reduced by a certain percentage.[61] Second, the sponsor may ask a non-risk-adjusted payment ("mandatory reinsurance premium") from the health plans[62]. Third, the sponsor may ask higher solidarity contributions from the consumers; at the aggregate level the additional contributions should equal, apart from transaction costs, the reductions in premium contributions that the health plans offer the consumers because of their reduced expenditures due to the risk sharing. The choice of how to finance the retrospective compensation may depend on the institutional context, the regulatory framework, the modality of the subsidy system (see Figure 1) and the precise form of the restrictions on the premium contributions. When introducing risk sharing the sponsor should to change the weights in the subsidy formula to adjust the premium subsidies for the new financial risks that health plans bear.

4.1. Forms of risk sharing

4.1.1. Risk sharing for all members

Risk sharing between the sponsor and the health plans can take several forms. The sponsor's retrospective payments may depend on the plan's acceptable costs, which serve as the basis for setting the risk-adjusted premium subsidies. Because

[59] Because it is hard to define the acceptable cost level of a health plan, in practice some proxy is generally used.
[60] The major functions of traditional reinsurance are to protect a health plan against insolvency and to increase its financial capacity to underwrite coverage [Bovbjerg (1992)].
[61] The sponsor's retrospective payments to the health plans can be considered a second type of subsidy from the sponsor to the health plans (see section 1.2).
[62] In this case the sponsor's role with respect to risk-sharing could be taken over by an independent insurer entity.

it is hard to define a plan's acceptable cost level, in practice the sponsor's retrospective payments often depend on the plan's actual incurred expenses or its imputed spending based on a price or fee schedule applied to observed utilization. In the latter case the incentives to produce the units of service efficiently are preserved. Further, the sponsor's retrospective payments may depend in different ways and to various degrees on the plan's acceptable cost or its proxy. For example, the sponsor may retrospectively reimburse each health plan a fixed percentage, e.g., 50%, of all its acceptable costs. This type of risk sharing has been proposed by Ellis and McGuire (1986, 1993); Gruenberg et al. (1986); and Newhouse (1986, 1994). They variously referred to it as "supply-side cost sharing", "partial capitation" and " a blend of capitation and fee-for-service". They discussed risk sharing in the context of modality A of the subsidy system (see Figure 1) with community-rated premium contributions. For the general case of any form of restrictions on the premium contributions and any modality of risk-adjusted subsidies we will refer to this type of risk sharing as *proportional risk sharing.* In the US, the widespread practice of experience rating health premiums at the employer (i.e. sponsor) level is a form of proportional risk sharing.

Another form of risk sharing is for the sponsor to compensate each health plan only for a certain percentage of the acceptable expenditures above a certain annual threshold, for example $ 20,000, per member.[63] Generally speaking, we will refer to this as *outlier risk sharing* .[64] There are clear analogies between outlier risk sharing (between the sponsor and the health plans) and the outlier pools for hospital and physician reimbursement [Keeler et al. (1988); Ellis and McGuire (1988)]. Outlier risk sharing requires that health plans account for all acceptable expenditures for each of their members. For the time being this requirement may reduce its practical applicability in some countries.

Risk sharing reduces both the incentive to deter nonpreferred risks and the incentive to attract preferred risks. The predicted losses on nonpreferred risk are reduced because the retrospective payments that a health plan expects to receive for persons who are above-average-risks within their premium-risk-group generally exceed their contribution to finance the sponsor's retrospective payments. Because the opposite holds for the low risks within each premium-risk-group, their predicted profits are reduced by risk sharing.[65]

4.1.2. Risk sharing for high-risks

A common feature of proportional and outlier risk sharing is that the health plan retrospectively receives compensation also for members whom it ex-ante did not consider high-risks, e.g., healthy persons who had a car accident. The retrospective compensation from the sponsor for those members does not contribute to reduce the *predictable* losses and therefore does not reduce the incentives for selection; it only reduces the health plan's incentive for efficiency. To improve the effectiveness of risk sharing Van de Ven and Van Vliet (1992, p. 38) proposed that a health plan itself decides for which members it will share the risk with the sponsor. According to their proposal each health plan would be allowed to *ex ante* designate a specified percentage of its members (for example, 1 or 4 percent) for whom the sponsor retrospectively would reimburse all or some acceptable expenditures[66]. In advance of the contract period (e.g., a year) each health plan would inform the sponsor which of its members it will share the risk with the sponsor. The group of selected members may change every contract period. A rational health plan will assign those members for risk sharing whom it predicts will have the highest losses. The risk sharing for these high-risk members could apply to a certain percentage of their expenses, or to their expenditures above a threshold, or to a combination. *"Risk sharing for high-risks"* is an effective tool to reduce the incentives for selection if health plans can predict *very high* losses for a *small* group of (potential) members, e.g., because they know the results of lab tests or genetic testing, or they know some specific medical conditions not accounted for by the risk adjusters.

---

[63] Another variant is that the sponsor compensates each health plan for a certain percentage of all normative expenditures above a *cumulative* annual threshold for *all* its members together. We refer to this as *stop-loss* risk-sharing. Although a stop-loss risk-sharing arrangement would provide the health plans with good solvency protection, the effect on the reduction of selection will probably be low. Only if a health plan expects its future annual expenditures to exceed the stop-loss limit is there no incentive for selection. Probably, however, the sponsor would want health plans to have an incentive for efficiency and therefore will set the stop-loss limit at such a level that the majority of the health plans do not exceed it.

[64] Beebe (1992) referred to it as outlier-pooling.

[65] These effects hold on average. The precise effect in an individual case depends on the form of risk sharing, the level of the predictable profit/loss without risk sharing, and the way that the sponsor's retrospective payments are financed.

[66] An alternative is that each health plan is allowed to *ex post* designate a specified percentage of its members, e.g., 1%, for whom the sponsor retrospectively reimburses all or some normative expenditures [Van de Ven et al. (1994, p.130)]. For a statistical analysis of this alternative, including a comparison with 'risk sharing for high-risks', see Van Vliet (1999). Because under this alternative the selected members would be all persons with losses above a certain threshold whose value a health plan may accurately predict, such an alternative is similar in its incentives for efficiency to a system of outlier risk-sharing. A difference is that the threshold amount is not the same for all health plans. Another difference concerns the incentives for selection [Van Barneveld, (1999)].

"Risk sharing for high-risks" can be considered a form of pro-competitive arrangement that, from the health plans' point of view, tries to simulate the free competitive market. A free health plan market may lead to discontinuity of coverage, through the refusal to insure some high-risk applicants or to the exclusion of pre-existing medical conditions [Light (1992)]. Instead of terminating the contract or refusing high-risk applicants, a health plan can now assign high-risk persons for risk sharing with the sponsor. This risk sharing arrangement can be organized such that the high-risk persons themselves are not aware of the risk sharing.

To improve the effectiveness of the risk sharing an information system could be set up to reduce the market imperfection that exists in case a health plan cannot accurately assess the financial risk that a new applicant generates [Newhouse (1994)]. Health plans could receive relevant (standardized) information from the prior plan, or from the sponsor (e.g., whether or not the person was selected for risk sharing in the prior year). Alternatively, a health plan could be allowed to have a health interview with newly enrolled members.

An advantage of "risk sharing for high-risks" is that it may reduce the health plan information surplus vis à vis the sponsor. Because health plans use their information surplus for selecting the high-risk applicants, an actuarial analysis of the risk-profile of the assigned high-risk members, when compared with that of non-assigned members, may provide the sponsor with useful information which can improve the risk adjustment mechanism in successive years. In this way the sponsor can progress in its attempts to incorporate in the risk adjustment mechanism as much information as the health plans have.

A problem with "risk sharing for high-risks", that does not occur with the other forms of risk sharing, is setting the premium contribution for the high-risk members. Without risk sharing a health plan will ask high-risk applicants to pay the *maximum* premium contribution that is allowed under the regulation. On the other hand, if the expenses for a high-risk member are retrospectively reimbursed by the sponsor, the appropriate premium contribution would be the *minimum* premium contribution allowed under the regulation. Because a health plan selects its members to be assigned for risk sharing after it knows all the members in its portfolio for the next contract period, the question is, which premium contribution should ex-ante be offered to a high-risk applicant: the maximum or the minimum premium contribution? The extent of this problem diminishes, of course, as the difference between the maximum and minimum premium contribution narrows. In the extreme case of community rated contributions all members per health plan should be quoted the same premium contribution.

The concept of "risk sharing for high-risks" has been studied by Van Barneveld et al. (1996, 1998) for modality A (see section 2) of the subsidy system with community-rated premium contributions within each health plan.[67] Van Barneveld et al. (1996) suggest that another variant in the case of poor risk adjusters is to have the percentage of members to be selected for risk sharing to depend on a health plan's average loss per member - before risk sharing and adjusted for the difference between the health plan's premium contribution and the national average premium contribution - in a preceding year. The rationale is that, especially with crude risk adjusters, these losses are caused mainly by the inability of the crude risk-adjusted capitation to compensate adequately for health status.

When discussing Medicare's method for reimbursing at-risk managed care plans in the US, Newhouse et al. (1997) proposed to implement "risk sharing for high-risks" in addition to proportional risk sharing for all members. They expect "risk sharing for high-risks" to be especially useful for dealing with the terminally ill.

4.1.3. Condition-specific risk sharing

So far we have discussed three forms of cost-based risk sharing. An alternative is to retrospectively reimburse the health plans some *prospectively* determined payments dependent on the occurrence of some medical problems (Luft, 1986; Enthoven, 1988). We refer to these arrangements as "*condition-specific risk sharing*".[68] The payments can be based

[67] Van Barnveld et al. (1996, 1998) refer to it as mandatory high-risk pooling. Major differences with the high-risk pools in the US [see Bovbjerg and Koller, (1986); Zellner, Haugen and Dowd (1993)] are that under mandatory high-risk pooling the high-risk members pay the same (community-rated) premium as others, they have the same benefits package and the same copayment-structure as others, and they are unaware that their health plan shares their risk with the sponsor.

[68] Condition-specific risk sharing differs subtly from retrospective risk adjustment (section 3.2.2) because with the latter the weights or payments are *retrospectively* determined. In practice this difference will be negligible, so that condition-specific risk-sharing and retrospective risk adjustment come to the same thing.

on diagnoses that are relatively invulnerable to manipulation and for which high cost treatment is relatively non-discretionary. Because the amount of the payment is prospectively determined, and not dependent on a health plan's actual costs, condition-specific risk sharing does not change the plan's incentive to produce the units of service efficiently.[69] Given that the goal of risk sharing is to reduce the incentives for selection, condition-specific risk sharing contributes to this goal only insofar as health plans ex ante know that there are individuals with an above average probability within their premium-risk-group of having or developing the specific condition, or as far as consumers ex ante know they have an above average probability within their premium-risk group to do so. In that case condition-specific risk sharing may be a valuable addition to the DCGs developed by Ash et al. (1989) and the DCG/HCCs developed by Ellis et al. (1996a). Otherwise, condition-specific risk sharing, just like traditional reinsurance, only reduces the health plan's financial risk, without reducing the incentives for selection.

An advantage of "risk sharing for high-risks" over condition-specific risk sharing is that it prevents "diagnosis-inflation" and political battles over which conditions are to be compensated [Swartz (1995)], and it does not retrospectively reimburse expenses for members whom the health plan *ex ante* did not consider to be high-risk persons. An advantage of condition-specific risk sharing over " risk sharing for high-risks" with a fixed percentage of the selected members is that plans that specialize in certain high-cost treatments and that are therefore flooded by high-risk members (because of adverse selection), receive appropriate compensation for *each* of these high-risk members. Health plans who have no high-risk members at all (possibly because of selection), do not receive any retrospective compensation.

4.2. Empirical results

In principle there exist many forms of risk sharing between the sponsor and the health plans. In this section we discuss some empirical results with respect to forms of risk sharing which have been reported in the literature. As illustrated in Table 4, many of these forms of risk sharing can be described by three parameters [Van Vliet (1997)]:

- r, the reimbursement rate;
- T, the threshold amount;
- p, the percentage of members, to be ex-ante assigned, to whom the risk sharing applies.

Table 4

There is no common terminology in the literature. Different authors use different terms for the variants of risk sharing. Here we adopt the terminology given in Table 4 [70].

The first empirical study, to our knowledge, on risk sharing in the context of risk adjustment is Beebe's (1992) analysis of outlier risk sharing, which he called an outlier pool. The pool would pay 45 percent of the expenditures above a threshold amount. Beebe varied the threshold between $ 100,000 and $ 10,000 (1992- US-dollars). He analyzed US Medicare data, so most of the sample were above the age of 65. The percentage of persons exceeding the threshold varied from 0.07 to 11.1 percent. The pool's payments varied from 0.14 to 19.5 percent of the total expenditures. Beebe concluded that an outlier pool payment method could provide some protection against the risk of an unexpectedly high proportion of high-cost users at a relatively modest cost. He did not examine the outlier pool's ability to reduce the incentives for selection.

Van Barneveld et al. (1996) analyzed the effects of "risk sharing for high-risks" for modality A of the subsidy system with age/gender-adjusted premium subsidies and with community-rated premium contributions. Their findings indicate that under these conditions the mean per capita loss for the 1 per cent of individuals with the highest prior-year expenditures are 8.5 times the overall mean per capita expenditures. This illustrates the great potential of " risk sharing for high-risks" to reduce the health plans' incentives for selection, because health plans could select their members simply on the basis of the prior year's costs. This appears to be an effective selection strategy[71]. To the extent that health plans are able to better predict which of their members belong to the long right

[69] Related to condition-specific risk-sharing is Beck and Zweifel's (1998) proposal to give health plans retrospectively a prospectively determined payment for each member who dies.
[70] An alternative is to adopt a common term, e.g., Risk-Sharing, and to specify the value of the above three parameters, r, T and p.
[71] If health plans would select their members for "risk-sharing of high-risks" on the basis of available information on prior hospitalizations and prior costs in the three preceding years, the mean per capita predictable loss for the 1% individuals with the highest losses would increase by less than 10 percent [Van Barneveld et al. (1998, Table 2)].

tail in the distribution of residual costs not accounted for by the risk adjusters used by the sponsor, the more effective is "risk sharing for high-risks". For the next 1 percent group with highest prior-year costs, the mean per capita predictable loss falls to 4.5 times the overall mean per capita expenditures. For the next two percentiles this ratio falls to about 2.5, and after that it falls below 1.5. From these figures Van Barneveld et al. (1996) conclude that risk sharing for less than 4 percent of the members would be most meaningful. If still more members were allowed to be selected, the marginal reduction of predictable losses would be small, while the incentives for efficiency would be lowered further.

Another illustration of the effectiveness of "risk sharing for high-risks" is that with age/gender adjusted capitation the mean per capita predictable loss for the 8 percent of individuals who were hospitalized *two years ago* is about 1.1 times the overall mean per capita expenditures. If the sponsor would allow risk sharing for 4 percent of the members, and if health plans would select these 4 percent members on the basis of *prior-year* expenditures, the predictable loss on those members who have been hospitalized two years ago, would be reduced by about two thirds [Van Barneveld et al. (1996, Table 3)]. That is, the gross returns on potential selection activities based on hospitalizations two years ago would be reduced by two thirds. Because of the costs of selection strategies, the net returns would go down even more. This reduction of the incentive for selection should not come at the expense of substantially reduced incentives for efficiency because the health plan remains fully at risk for the 75 percent of the total expenditures caused by the 96 percent non-selected members. Further, because a health plan remains responsible for the high costs of persons with *unpredictable* high expenditures, which comprise the majority of all high costs, the selected members may be "free riders" as far as the health plan's managed care activities are concerned. If the sponsor turns out to bear the major financial responsibility for certain medical treatment programs, e.g., transplantation, open heart surgery or HIV-treatment, the sponsor should be involved in the process of managing these types of care.

In another study Van Barneveld et al. (1998) compared, under the same conditions as above, the effectiveness of proportional risk sharing, outlier risk sharing with 100 per cent reimbursement and "risk sharing for high-risks" with 100 per cent reimbursement and no threshold. Because of the tradeoff between efficiency and selection, they chose the values of the parameters r in proportional risk sharing, T in outlier risk sharing and p in 'risk sharing for high-risk" such that on average the percentage of the total costs for which the health plans are at risk, was the same for each form of risk sharing. Based on the prior discussion, an approximately optimal choice of p appeared to be 4 percent. The corresponding values for r and T were 20 percent and 10 times the mean spending respectively. For each form of risk sharing the premium subsidies were proportionately reduced to keep the sponsor's total outlay the same. (This reduces the incentive to attract good risks.) As an indicator of the effectiveness of the different forms of risk sharing they used the reduction of the mean per capita predictable loss for the 20 percent of individuals who had been hospitalized in the four preceding years. Without risk sharing this predictable loss is slightly more than the overall mean per capita expenditures. Proportional risk sharing reduced this predictable loss by 20 percent, outlier risk sharing reduced it by 41 percent and "risk sharing for high-risk" by 51 percent. The predictable profits on the 80 percent individuals who had not been hospitalized in the four preceding years, are reduced by the same percentages. Therefore, Van Barneveld et al. (1998) conclude that "risk sharing for high-risks" is more effective in reducing the incentive for selection than the two other forms of risk sharing.

Van Vliet (1997) concluded that the effectiveness (in terms of reducing the plans' incentive for selection) of "risk sharing for high-risks", relative to proportional and outlier risk sharing, can be further increased by reducing the reimbursement rate, e.g.,. from 100 percent to 80 percent, while at the same time increasing the percentage of selected members (keeping total retrospective payments constant).

A different type of empirical study has been done by Keeler et al. (1998). They simulated the effect of several forms of risk sharing on the adverse selection that occurs if consumers have an annual choice among three different health plans with varying generosity of coverage. The three simulated plans differed only in the cost-effectiveness ratio that their treatments should surpass. The expenses of the generous plan are nearly double the expenses of the stingy plan for an average case-mix population. It is assumed that health plans cannot use treatment policy to discriminate against the sick. The sponsor requires the plans to ask a community-rated premium contribution from their members (modality A of the subsidy system). Further it is assumed that health plans are not actively selecting healthy enrollees by other forms of selection than the differentiation of the generosity of their coverage. Consumers with different health status, income and tastes for health care are assumed to choose their most preferred health plan during the annual open enrollment period. The acceptable costs, on which the sponsor's subsidy (capitation) is

based, are the costs of the middle plan. A first conclusion from this simulation is that flat capitation, i.e. no risk adjustment at all, results in severe adverse selection. The healthy individuals are overrepresented in the stingy plan and the sick in the generous plan. Without the assumption that half the persons will stay in their original plan, there would be no equilibrium. Because the sponsor's subsidy is the same for each consumer, the stingy plan appears to be overcompensated by 30 percent, relative to its risk-profile, and the generous plan undercompensated by 37 percent. Keeler et al. (1998) simulated the effects of several forms of risk sharing on the extent of the sponsor's overpaying and underpaying. They found that proportional risk sharing at a 25 percent reimbursement rate and outlier risk sharing of in total 10 percent of the sponsor's outlay, each reduced the sponsor's over- and underpaying by 35 to 50 percent. A form of condition-specific risk sharing which compensates about 25 percent of the plans' expenses, reduced the sponsor's over- and underpaying by about two thirds. With all forms of risk sharing the capitation payments are proportionately reduced to keep the sponsor's total outlay the same.

4.3. Discussion

4.3.1. What form of risk sharing is optimal?

Given the effectiveness of different forms of risk sharing to reduce selection Van Barneveld et al. (1998) conclude that "risk sharing for high-risks" should be preferred rather than proportional or outlier risk sharing. This conclusion seemingly conflicts with the view of Newhouse et al. (1997) who argue for proportional risk sharing rather than for other forms of risk sharing. The explanation for these seemingly different conclusions is that Van Barneveld et al. consider risk sharing only as a tool to reduce the incentives for selection in case of imperfect risk adjustment with restrictions on the premium contributions, while Newhouse et al. consider it also as a tool to reduce the incentive for quality skimping, which may occur even in the absence of any incentive for selection.

By *quality skimping* we mean the reduction of the quality of care to a level which is below the minimum level that is acceptable to society. The argument is that if a health plan's marginal revenue is zero for the additional services that its members receive, the plan may have a incentive for quality skimping. Although perfect information and competition in the plan market would prevent underprovision, one should not exclude the possibility that the same information problems that enabled fee-for-service providers to order and profit from excess care, may prevent patients or their agents from punishing underprovision [Keeler et al. (1998)]. According to Newhouse et al. (1997) the ideal form of risk sharing pays a plan a prospectively set marginal cost and a capitation such that the plan breaks even on that case. This would address both concerns of selection and quality skimping[72]. In practice, however, we do not have marginal cost and thus will not have an ideal payment system[73].

The extent to which risk sharing can be an effective tool to reduce incentives for quality skimping, even in the absence of any incentive for cream skimming, depends (at least) on the type of health plan.[74] We can discern several types of health plans. Our definition of health plan (see section 1) is "a risk-bearing entity that performs at least some insurance function, but that may also manage or provide health care". So, on the one hand, a health plan can be a traditional commercial insurer that has no contractual relationship with the providers of care and which (partly) reimburses the fee-for-service bills sent by the providers to the consumers [what Enthoven (1994) calls a "remote third-party payer"]; and on the other hand a health plan can be a managed care organization, e.g., a capitated provider group, that itself delivers the care to its members. Only for the latter type of health plan is the literature on optimal provider reimbursement relevant. In this literature it is argued that , rather than full (risk-adjusted) capitation, some form of reduced supply-side cost-sharing is optimal [Ellis and McGuire (1986, 1993); see also Pauly (1980)]. In our terminology this could be a form of risk sharing between the sponsor and the health plans[75]. However, if the health plan is a traditional "remote third-party payer", the arguments about optimal provider reimbursement do not influence the optimal structure of the insurance premium [Ellis and McGuire (1990); Selden (1990)] and risk sharing between the sponsor and the health plans cannot be considered a tool for reducing the incentives for quality skimping.

[72] Concerns of moral hazard remain [Newhouse (1986)].
[73] For an extensive discussion on the relation between capitation and quality of care, see the chapters on physician payment by Tom McGuire and by Mark Pauly (this Handbook).
[74] It may also depends on the type of benefits included in the health plan coverage [see Van de Ven and Schut (1994)].
[75] An alternative is direct consumer cost-sharing with a coinsurance rate of e.g., 30% [Manning and Marquis (1996)] or prior-cost as a rating factor in the premium contribution model with a weight of e.g., 0.30.

4.3.2. Proportional risk sharing or prior costs as a risk adjuster?

So far we have considered risk sharing as complementary to insufficient risk adjustment. So, we discussed risk adjustment and risk sharing as separated issues. However, there is a similarity between the two, as Newhouse (1986) argued. He compared the situation where the sponsor subsidy to the health plans depends on prior cost (or prior utilization) with the situation that it depends on current cost (or current utilization). Prior costs is used as a risk adjuster in the premium subsidy formula. Current cost is applied in the form of a blend of capitation (not dependent on prior costs) and current costs. We refer to the latter as proportional risk sharing. Newhouse (1986) argued that prior cost and current cost are similar in their incentive effects, or can be made so (except for those who die or switch plans)[76]. Given this similarity, Newhouse favors proportional risk sharing rather than prior-costs-adjusted capitation because current utilization shows recognition of changes in health status as they occur, rather than with a delay. Newhouse considers current utilization a more sensitive measure of predictable variation in expected cost then prior utilization. [77]

Some other points can be noted if the sponsor subsidy depends on either prior costs or current costs (or utilization). First, the way that the retrospective payments from the sponsor in the case of proportional risk sharing are financed (see above) may have distributional effects that differ from those of prior-cost-adjusted subsidies. Second, the weights given to the other adjusters may change after inclusion of prior costs in the subsidy formula. Third, the premium subsidy need not be a linear function of prior costs, and it may depend on several years claims records rather than one year's claims records. Fourth, prior costs or prior utilization as a risk adjuster may need some adjustment in case of the opening/closing of hospitals and other health care facilities in the region.

4.3.3. Ongoing research

The research on risk sharing as a tool to reduce the incentives for selection is in an early stage. Because risk adjustment in principle is the preferred option to prevent selection, it is not surprising that the research on risk sharing started some ten years later than the risk adjustment research. Given the growing consensus in the literature about the need for some form of risk sharing to complement imperfect risk adjustment, and given the primitive forms of risk adjustment currently used in most venues (see section 5), and given the growing awareness that "it is now clear that risk adjustment is a very complex technical issue, and that it will be extremely expensive to try to build the capability to create close to perfect risk adjusters" [Swartz (1995)], there is a growing need for further research on risk sharing[78].

Future research should focus on getting to know the terms of the tradeoff between efficiency and selection. A conceptual framework could be built to weigh efficiency against selection, taking into account the different types of managed care strategies, e.g., case by case management versus the management of special treatment programs, as well as the different types of selection strategies as mentioned in section 2.4. Prediction models should be developed that health plans could use in practice, given the information in their administration. Future research should then try to answer questions like: What is the distribution of predictable profits and losses if health plans use their best prediction model, given a certain subsidy formula and given certain restrictions on the premium contribution? How do we value an overall reduction of the predictable losses versus a selective reduction of the highest predictable losses only? What are the optimal values of the parameters of risk sharing for several subsidy-formulae and several forms of restrictions on the premium contribution? Future research could also focus on the consequences of risk sharing on the subsidy formula, i.e. the recalculation of the premium subsidies, and on the health plan's premium setting. Finally, attention should be paid to the similarities and differences between forms of risk sharing (proportional and outlier risk sharing) and prior costs as a risk adjuster.

[76] The same similarity exists between outlier-risk-sharing and a risk-adjuster "high prior cost", i.e. prior costs only as far they exceed a certain threshold, e.g., the 99 percentile of the empirical distribution of costs [as applied by Lamers and Van Vliet (1996)].
[77] Another argument is that in the Medicare system in the U.S. no prior cost (or utilization) data are available for a new cohort of enrollees.
[78] For some work in progress concerning risk-sharing, see Van Barneveld (1999).

5. The practice of risk adjustment and risk sharing

5.1. International comparison

In the late 1990s risk adjustment and risk sharing are being applied in competitive health plan markets in at least eleven countries (see Table 5 and 6)[79]. All countries, except the US, implemented these financing mechanisms in the 1990s. In all countries, expect Ireland and Switzerland, the solidarity contributions are income-related. Most countries use age and gender as risk adjusters. In addition, some countries adjust the sponsor premium subsidy also for region and disability. The most predictive risk adjusters mentioned in section 3, are not yet in common use. The major exception is the US, where some programs have implemented diagnosis-based risk adjustment [Dunn (1998)] and where the Medicare program has announced that it will use diagnosis-based risk adjustment in the year 2000 to pay HMOs for their enrollees [Greenwald et al. (1998)]. We speculate that three major barriers have contributed to the delayed implementation of risk adjustment in many countries: the recency of the most promising research, the non-availability of relevant data, and inertia. Because in most countries risk adjustment is a dynamic process[80], over time we may expect to see research results to be implemented in practice.

Table 5 and 6

All countries listed in Table 5 and 6 have stringent restrictions on the variation of premium contributions. In four, the sponsor requires the premium contribution to be zero, that is, the premium must equal the premium subsidy. In the other seven countries the health plans are allowed to ask from their members a community-rated premium contribution. As discussed in section 2 and 3 the combination of poor risk adjusters and stringent restrictions on the premium contributions results in large predictable losses on high-risk individuals. Given this conclusion it is surprising to see that one half of the countries mentioned in Table 5 and 6 have no form of risk sharing to reduce the predictable losses and profits.

Despite the strong financial incentives created by capitation payments, selection and its adverse effects have only infrequently been reported as a major problem in most countries[81]. The primary example where selection is pervasive is the US, where there is considerable evidence that health maintenance organizations (HMOs) enjoy favorable selection[82] and where concern is growing about the adverse effects of selection on the quality of care, especially for high-risk patients. In its recent Report to Congress, on Medicare Payment policy, the Medicare Payment Advisory Commission [MEDPAC (1998)] highlights that new enrollees in Medicare managed care plans cost about 35 percent less than the Medicare fee-for-service average in the six months prior to enrollment. In contrast, Medicare expenditures on persons disenrolling from HMOs averaged 60 percent above average in the six months following disenrollment. This finding is also supported by comparisons using self-reported health status measures from the 1994 Medicare Current Beneficiary Survey [Riley et al. (1996)].

Several arguments explain why selection may not be a major issue in the early stage of the implementation of a risk adjustment mechanism in a (potentially) competitive health plan market, and why over time selection may increasingly become a problem. First, in the early stage many players may be unfamiliar with the rules of the game. For example, in the Netherlands, even five years after the implementation of the consumer's right to change health plans annually, many consumers were unaware of this right. In addition, people often associate changing health plans with a potential non-acceptance, the exclusion of pre-existing medical conditions, or higher premiums. Also in the early stage not all managers working within in a health plan fully understand the financial incentives of the financing mechanism. So, in the beginning this lack of knowledge, which is enlarged by the complexity of the risk adjustment system and by the annual changes in the system, may restrict the selection problems. However, over time consumers and managers will be better informed and can be expected to react to large incentives for selection that

[79] In other countries proposals for a competitive health plan market are under discussion, e.g., Poland (to be implemented in 1999), Argentina, Chile, Portugal and Taiwan.

[80] For example, in the Netherlands and the United Kingdom in 1991 and 1992 the subsidy was based on (estimated) prior costs at the plan level. Subsequently more risk-adjusters were implemented.

[81] For an academic discussion on the potential threat of cream skimming, see e.g., Newhouse (1982), Luft (1986), Van de Ven and Van Vliet (1992) and Matsaganis and Glennerster (1994). This discussion is summarized in section 2.

[82] See e.g., Hellinger (1995), Lichtenstein et al. (1991), Luft and Miller (1988) and Robinson et al. (1993).

occur in the system without risk adjustment. Secondly, in the early stage in most countries the differences among health plans with respect to benefits package, premium contribution and contracted providers are relatively small. Over time, especially when less stringent government regulation with respect to planning facilities and medical pricing permits health plans to diversify the conditions of the contracts with their members, we may see more market segmentation . Thirdly, in most countries the risk adjustment mechanism has been implemented in the (mandatory) social health insurance sector. Traditionally most of the health plans working in that sector are highly driven by social motives rather than by financial incentives. However, with open entry for new health plans (subject to certain conditions), as is the case in all countries except Belgium, new health plans may make the behavior of the traditional health plans more incentive driven. As the chief-executive-officer of a large Dutch sickness fund said: We are administering the social health insurance now with one additional limiting condition, i.e. "our expenses should not exceed our revenues"[83]. Finally, one may argue that selection is not so much of a problem because doctors may be reluctant to discriminate among risks because of medical ethics. However, present ethics may change if the entire delivery system becomes more competitive. We share Newhouse's (1982) skepticism that medical ethics are sufficient to make selection effects unimportant. In our opinion, appropriate financial incentives and appropriate rules of the game should provide the ultimate safeguard against the adverse effects of selection.

From Tables 5 and 6 we see that a variety of forms of risk sharing are applied in practice. In Belgium (1998) the health plans are at risk for about 3 percent of their expenses[84]. In the Netherlands (1998) proportional risk sharing (with r=0.95) for the fixed (i.e. production independent) hospital costs[85] (about one third of the total health care expenses) is combined with proportional risk sharing (r=0.40) and outlier risk sharing (with r=0.90 and T= 4500 guilders) for all other health care expenses[86]. New York State applies a form of condition-specific risk sharing in its Medicaid program. In the United Kingdom (UK) the general practitioners (GP) fundholders, who in our terminology can be considered health plans, are fully compensated for all expenses above £ 6000 per person per year. The fundholders' budgets are determined by negotiations around an "activity-based capitation bench-mark" based on the age/gender characteristics of the practice, and also the practice's historic activity [McCarthy et al. (1995); Sheldon et al. (1994)]. Because negotiations allow the influence of other local factors or expenditure to be included in the budgets, some implicit forms of risk sharing are already incorporated into the budgets.

Most countries have an annual open enrollment period. Two major exceptions are the US and the UK.[87] Until 2002 Medicare members in the US may change health plans every month, which gives more opportunities for selection than does annual choice. In the UK the GP-fundholders may refuse to accept new patients and they have the right to request that any patient should be removed from their list without explanation [McCarthy et al. (1995)]. In 1992-93 78,000 patients, about one in 600 of the population of England, were removed from a GP's list specifically at the request of the GP [Bevan and Sheldon (1996)].

In seven countries membership in a capitated health plan is a mandatory aspect of the social health insurance system. In three countries consumers have alternative options within the system: they may choose the traditional public system (Colombia and UK) or the traditional fee-for-service system (US Medicare). In Ireland the risk adjustment mechanism applies to voluntary private health insurance, which is complementary to the National Health Service. Although the whole Irish population is entitled to receive public health care, about one third of the population has a voluntary supplementary insurance, primarily to receive private care and thus bypass the queue in the public system. The various selection effects (see Table 1) may depend on *voluntary* or *mandatory* membership in a capitated health plan. On the one hand the voluntary character of supplementary insurance with community-rated consumer contributions in Ireland speeds up adverse selection. The low risks simply do not buy the insurance (and thereby do not pay a solidarity contribution), resulting in a continuously upward premium spiral. On the other hand, the voluntary character of membership in a capitated health plan in the US Medicare system may dampen some of the adverse effects of selection. With mandatory membership high-risk consumers would be forced to choose one of the competing capitated health plans, each of which has a disincentive to be responsive to their preferences. This could potentially result in poor service and poor quality of care for them. Due to the voluntary character of membership in

[83] Until 1991 the Dutch sickness funds received from the sponsor a full reimbursement of all their expenses.
[84] Another unique aspect of the Belgium system is that the benefits package for which the health plans are at risk, also covers long term care. This may be an (additional) argument for not giving the health plans too much financial risk (see Van de Ven and Schut, 1994).
[85] The subsidy for the fixed hospital costs is based on a plan's prior year costs.
[86] The threshold of 4500 guilders is slightly more than double the average per capita total annual expenditures. In 1999 the threshold went up to 7.500 guilders.
[87] Another exception is Germany, where the firm-based sickness funds are exempted from the open enrollment requirement.

a capitated plan, the high-risk persons are able to choose, as an alternative, the traditional fee-for-service sector, where physician fees likely exceed their marginal costs.

In the US risk adjustment is applied by different types of sponsors (federal government, states, employer groups). In addition to the projects mentioned in Table 6 the Ohio Medicaid program applies risk adjustment and risk sharing with contracted health plans [Kronick et al. (1995, p.20)], and risk adjustment programs using diagnosis-based information have been implemented in the late 1990s in Washington, Minnesota, and Colorado [(Lee and Rogal (1997); Dunn (1998)].

In addition to the above mentioned differences in risk adjusters and forms of risk sharing, there also appear to be a great variety in the number of competing health plans, the modality of the subsidy system (A or B), and the institutional context and regulatory regime. For example, there are large differences in the extent to which health plans are allowed to manage the care, e.g., by selective contracting , by negotiating prices, by building new facilities, or by buying new equipment. There are also differences in the conditions to be fulfilled by new health plans entering the market and differences in the benefits to be covered. Because of all these differences it is hard to compare and evaluate the effects of different forms of risk adjustment and risk sharing in practice.

5.2. Problems in practice

A major practical problem is the availability of data with which to risk adjust. In some countries (e.g., Belgium, Colombia, Germany, Israel and Russia) there are no data available that link individual consumer characteristics with individual health care expenditures. So the first generation risk adjusters in these countries are based on available aggregated data. Some subsidy formulae are based on utilization data per age/gender group for major types of care, weighted by their relative proportion in total health care spending[88]. Over time, as individual expenditure data become available, better subsidy formulae can be applied[89]. Another problem is that for some risk adjusters the average per capita expenditures are known only per subgroup of this risk adjuster, but not for the sub-subgroups in interaction with the other risk adjusters. Consequently some sub-subgroups are over - or undercompensated because of correlations between risk adjusters.

The (non-)availability of data largely influences the type of risk adjusters or risk sharing to be used. For example, the application of DCGs or HCCs as risk adjusters requires that the sponsor has access to the relevant diagnostic information of the members of each health plan. In the US Medicare system this is less of a problem than elsewhere (Ellis et al. 1996b). For example, in the Netherlands the specialists working in the hospital refuse, for privacy reasons, to provide the sickness funds with diagnostic information about individual hospital admissions. On the other hand, in the Netherlands detailed information on health care expenditures (also per subsidy-risk-group) per sickness fund is routinely available to the sponsor. As a result the marginal administrative costs of risk sharing between the sponsor and the sickness funds are relatively low. Because these cost data are not routinely available in HMOs in the US, Beebe (1992) proposed to use hospital stays as the basis for outlier risk sharing. So, the availability of either diagnostic information or prior costs may influence the sponsor's preference for either DCGs/HCCs or prior costs as a risk adjuster, as well as for either condition-specific risk sharing or some other form of risk sharing.

A practical problem with risk sharing is assessing the acceptable costs that form the basis for risk sharing and for setting the risk-adjusted premium subsidies. In the Netherlands there is a very detailed specification of the basic benefits package in combination with a standardized fee schedule and the sponsor closely monitors all the sickness funds' expenditures and decides which expenses are acceptable and which are not. This procedure will become more complicated the more degrees of freedom the plans have for managing the care and negotiating different price- and quality-levels, and the more the health plans integrate the specified basic benefits coverage with supplementary health insurance (as HMO's do in the US Medicare system). The problem of the acceptable cost-level is related to

[88] For the details of the subsidy formula and the risk-sharing arrangement in eight countries, see McCarthy et al. (1995), Kennedy (1996) and Schokkaert et al. (1996).

[89] For the development of the subsidy formula the sponsor should ideally have at its disposal a large data base with individual consumer information about expenditures and as many risk factors as possible. If the sponsor had the same information as the health plans routinely administer, the sponsor can simulate the predictable losses and profits for subgroups known to the health plans. For the day-to-day operational administration of the risk-adjusted subsidies it is sufficient if the sponsor knows the per capita normative expenditures per subsidy-risk-group and each health plan's number of members per subsidy-risk-group.

the lack of clinical consensus on the treatment of certain conditions. It is also related to the distinction between the S-type and N-type risk factors (see section 2.2.), i.e. the risk factors for which solidarity is desired or not desired. This distinction appears to be an issue especially in Belgium [Schokkaert et al. (1998)].

The problem of the non-availability of data for risk sharing may be reduced by using data that the health plans routinely collect for their own reinsurance. With respect to risk adjustment the sponsor may tackle the data problem by announcing that, after a reasonable period of time, higher subsidies for certain subgroups will be given only if the consumer or the health plan provides the sponsor with certain information. This provides the health plans with an incentive to routinely collect the required data.

A conclusion we can draw from the experience in practice is that even the simplest risk adjustment mechanisms are complex [Gauthier et al. (1995)] and that start up "surprise problems" can be expected [Bowen (1995)]. However, several sites in the US in the 1990s have made progress with the implementation of health based risk adjustment, suggesting that it is indeed possible to overcome both technical and political hurdles [Rogal and Gauthier (1998)].

6. Directions for future research

We have already covered so much that it would be very difficult to try to summarize it. Instead, we would like to end by identifying a few topics that have not yet received significant attention, but seem likely to be the focus of significant research in the future.

Risk adjustment has already come a long way over the past two decades, increasing both in its predictive power and in its sensitivity to creating appropriate incentives. It appears likely that the next decade will also see large improvements in predictive power, with the improvements coming in many areas. Those that seem most promising to us include using more refined clinical information (such as the results of lab tests or clinical chart information); pharmaceutical data; combining claims with self-reported information; or building better models of patterns of service use over time. Episode-based models, and models that make better use of the timing of new information generated during the year also hold out promise of improving the predictive power of risk adjustment models. Models that predict individual level expenditures on specific services instead of in aggregate terms may also hold out promise. We may also expect to see more realistic simulations of health plans' incentives for selection. This might entail weighting profits and losses unequally, for example, by giving greater weight to larger profits and losses than to small ones, or by giving greater weight to profits and losses that persist over time than those that occur only in the short term. In addition, models may be developed that compensate high-risk individuals for their above average fluctuations around the expected spending.

We have spent a considerable amount of effort in this chapter documenting the many diverse ways in which health plans may behave strategically in order to attract or retain profitable enrollees. Clearly as risk adjustment is implemented in more and more countries in more and more settings, it will be important to generate both theoretical models and empirical measures of the magnitudes of this behavioral response.

One reason for understanding health plan behavioral response to the premium subsidy calculations is to better inform the theory and empirical implementation of optimal risk adjustment research. Analogously to the extensive research of the last decade that has attempted to model and understand provider response to the method used to reimburse them, we anticipate that the next decade will see a proliferation of research that examines how premium payments, premium subsidies, and ex post risk sharing to health plans influences plan level behavior.

In order to understand how health plan behaviors will influence the enrollment patterns of consumers that choose among competing health plans, it would be useful to understand how well consumers can anticipate their own health care needs (e.g., using information from genetic testing), and how willing they are to change health plan enrollment in order to act on these expectations. Indeed there is a significant literature on individual choice of health plans, but greater attention could be paid to how these choice variables and consumer information are related to expected spending. A great deal of attention has focused on determining how much of the variation in health care spending is potentially explainable using individual level information. Yet it may turn out that consumers are either more naive than the researcher's predictive models, or consumer inertia or noneconomic factors result in selection problems that are less serious than predicted by the models.

We have highlighted that regulation of plan level competition and standardization of many plan features is an important mechanism for constraining cream skimming and other forms of selection activities. In addition to studying selection behaviors, studying the effectiveness of different regulations would be very helpful.

Risk adjustment and risk sharing are two different strategies for reducing risk selection incentives. Although there is a considerable literature on each, we are not aware of a literature that has examined the tradeoff between the two approaches or carefully examined how risk-sharing arrangements alter the desired risk adjustment formulas. This line of research seems particularly relevant given that in practice risk sharing arrangements are very common at the time that risk adjustment is introduced. In addition, future research could develop criteria for comparing different forms of risk sharing that aims at reducing both the incentives for selection and the incentives for quality skimping.

This paper has made a first step at assembling a few tables that compare risk adjustment and risk sharing internationally. Perhaps as interesting as studying settings where risk adjustment is being introduced is to understand why it has to date been so rarely used. As further experimentation goes on, it will also be helpful if countries could learn from the mistakes and successes of other countries. This would require that comparisons take place on a regular and systematic basis.

References

Altman, D., Cutler, D.M. and Zeckhauser, R. (1998) 'Adverse selection and adverse retention', *American Economic Review*, 88(2):122-126.

Anderson, G.F., Steinberg, E.P., Powe, N.R., Antebi, S., Whittle, J., Horn, S. and Herbert, R. (1990) 'Setting payment rates for capitated systems: a comparison of various alternatives', *Inquiry*, 27:225-233.

Arrow, K.J. (1963) 'Uncertainty and the welfare economics of medical care', *American Economic Review*, 53(5):940-973.

Ash, A. S. and Byrne-Logan, S. (1998) 'How well do models work? Predicting health care costs', in Proceedings of the Section on Statistics in Epidemiology, American Statistical Association, 42-49.

Ash, A., Ellis, R.P., Yu, W. et al. (1998) 'Risk adjusted payment models for the non-elderly', Final report for Health Care Financing Administration, June.

Ash, A., Porell, F., Gruenberg, L. et al. (1989) 'Adjusting Medicare capitation payments using prior hospitalization data', *Health Care Financing Review*,10(4):17-29.

Ashley, J. and McLachlan, W. (eds) (1985) *Mortal or morbid? - A diagnosis of the morbidity factor*. Nuffield Provincial Hospitals Trust, London.

Beck, K. and Zweifel, P. (1998) 'Cream-skimming in deregulated social health insurance: evidence from Switzerland', in P. Zweifel, ed., *Health, the medical profession and regulation*. Kluwer, Dordrecht, the Netherlands, 211-227.

Beebe, J.C. (1992) 'An outlier pool for Medicare HMO payments', *Health Care Financing Review*, 14:59-63.

Beebe, J.C., Lubitz, J. and Eggers, P. (1985) 'Using prior utilization to determine payments of Medicare enrollees in health maintenance organizations', *Health Care Financing Review*, 3(6):27-38.

Bevan, G. and Sheldon, T. (1996) 'Review of methods of risk rating in the UK National Health Service', in J. Hermesse, ed., *Risk structure compensation, Proceedings of an AIM workshop*. Maastricht.

Bovbjerg, R.R. (1992) 'Reform of financing for health coverage: what can reinsurance accomplish?', *Inquiry*, 29:158-175.

Bovbjerg, R.R. and Koller, C.F. (1986) 'State health insurance pools: current performance, future prospects', *Inquiry*, 23:111-121.

Bowen, B. (1995) 'The practice of risk adjustment', *Inquiry*, 32:33-40.

Bradshaw, J. (1972) 'A taxonomy of social need', in G. McLachlan, *Problems and progress in medical care*. Oxford University Press, London, 71-82.

Buchmueller, T.C. (1997) 'Managed competition in California's small-group insurance market', *Health Affairs*, 16(2):218-228.

Buchmueller, T.C. and Feldstein, P.J. (1997) 'The effect of price on switching among health plans', *Journal of Health Economics*, 16:231-247.

Butler, S.M. and Moffit, R.E. (1995) 'The FEHBP as a model for a new Medicare program', *Health Affairs*, 14 (Winter):47-61.

Carter, G., Newhouse, J. and Relles, D. (1990) 'How much change in the case mix index is DRG creep?', *Journal of Health Economics*, 9(4):411-428.

Carter, G.M., Bell, R.M., Dubois, R.W. et al. (1997) *A clinically detailed risk information system for cost.* Report for HCFA, RAND report number DRU-1731-1-HCFA November.

Chapman, J. D. (1997) 'Biased enrollment and risk adjustment for health plans', unpublished doctoral dissertation, Harvard University.

Chernichovsky, D. and Chinitz, D. (1995) 'The political economy of health system reform in Israel', *Health Economics* , 4:127-141.

Chinitz, D., Preker, A. and Wasem, J. (1998) 'Balancing competition and solidarity in health care financing' in R.B. Saltman, J. Figueras and C. Sakellarides, eds., *Critical Challenges for Health Care Reform in Europe*. Open University Press, Buckingham, U.K., 55-77.

Clark, D.O., von Korff, M., Saunders, K., Baluch, W.M. and Simon, G.E. (1995) 'A chronic disease score with empirically derived weights', *Medical Care*, 33(8):783-795.

Cochrane, J.H. (1995) 'Time-consistent health insurance', *Journal of Political Economy,* 103(3):445-473.

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

Diamond, P. (1992) 'Organizing the health insurance market', *Econometrica,* 60(6):1233-1254.

Duan, N. et al. (1983) 'Smearing estimate: a nonparametric retransformation method', *Journal of the American Statistical Association*, 78:605-690.

Dunn, D.L. (1998) 'Applications of health risk adjustment: What can be learned from experience to date?', *Inquiry*, 35:132-147.

Dunn, D. L., Rosenblatt, A., Tiaira, D. A. et al. (1996) *A comparative analysis of methods of health risk assessment*. Final report to the Society of Actuaries.

Efron, B. (1978) 'Regression and ANOVA with zero-one data: measures of residual variation', *Journal American Statistical Association* , 73:113-121.

Ellis, R.P. (1985) 'The effect of prior-year health expenditures on health coverage plan choice', in R.M. Scheffler and L.F. Rossiter, *Advances in Health Economics and Health Services*, Volume 6. Greenwich, CT, JAI Press, 247-272.

Ellis, R. P. (1990) *Time dependent DCG model*. Report prepared for Health Care Financing Administration, June, Boston University.

Ellis, R.P. (1998) 'Creaming, skimping, and dumping: provider competition on the intensive and extensive margins', *Journal of Health Economics*, 17:537-555.

Ellis, R. P. and Ash, A. (1989) *The continuous update Diagnostic Cost Group model*. Report prepared for the Health Care Financing Administration, June. Boston University.

Ellis, R.P. and Ash, A. (1995) 'Refinements to the Diagnostic Cost Group model', *Inquiry*, Winter:418-429.

Ellis R.P. and Azzone, V. (1998) 'OLS, loglinear and two part models of health expenditure: What do the data tell us?', Unpublished working paper, Boston University.

Ellis, R.P. and McGuire, T.G. (1986) 'Providers behavior under prospective reimbursement: cost sharing and supply', *Journal of Health Economics*, 5:129-151.

Ellis, R.P. and McGuire, T.G. (1988) 'Insurance principles and the design of prospective payment systems', *Journal of Health Economics*, 7:215-237.

Ellis, R.P. and McGuire, T.G. (1990) 'Optimal payment systems for health services', *Journal of Health Economics*, 9:375-396.

Ellis, R.P. and McGuire, T.G. (1993) 'Supply-side and demand-side cost sharing in health care', *Journal of Economic Perspective*, 7:135-151.

Ellis, R.P., Pope, G.C. , Iezzoni, L.I., Ayanian, J.Z., Bates, D.W., Burstin, H. and Ash, A.S. (1996a) 'Diagnosis-based risk adjustment for Medicare capitation payments', *Health Care Financing Review*, 12(3):101-128.

Ellis, R.P., Pope, G. C., Iezzoni, L. et al. (1996b) *Diagnostic Cost Group (DCG) and Hierarchical Coexisting Conditions and Procedures (HCCP) Models for Medicare risk adjustment*. Final report, Health Economics Research, for Health Care Financing Administration, April.

Encinosa, W. (1998) 'Risk adjusting in imperfectly competitive markets. Agency for health care policy research', unpublished paper.

Enthoven, A.C. (1978) 'Consumer-choice health plan', *New England Journal of Medi*cine, 650-658 and 709-720.

Enthoven, A.C. (1986) 'Managed competition in health care and the unfinished agenda', *Health Care Financing Review*, Annual Supplement, 105-120.

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

Enthoven, A.C. (1994) 'On the ideal market structure for third-party purchasing of health care', *Social Science and Medicine*, 39:1413-1424.

Epstein, A.M. and Cumella, E.J. (1988) 'Capitation payment: using predictors of medical utilization to adjust rates', *Health Care Financing Review*, 10(1):51-69.

Ettner, S.L., Frank, R.G., McGuire, T.G., Newhouse, J.P. and Notman, E.H. (1998) 'Risk adjustment of mental health and substance abuse payments', *Inquiry*, 35:223-239.

Farley, D.O., Carter, G.M., Kallich, J.D. et al. (1996) 'Modified capitation and treatment incentives for end stage renal disease', *Health Care Financing Review* , 17(3):129-142.

Files, A. and Murray, M. (1995) 'German risk structure compensation: enhancing equity and effectiveness', *Inquiry*, 32:300-309.

Fowler, E.J. and Anderson, G.F. (1996) 'Capitation adjustment for pediatric populations', *Pediatrics*, 98(1):10-17, July.

Fowles, J.B.,Weiner, J.P., Knutson, D., Fowler, E., Tucker, A.M. and Ireland, M. (1996) 'Taking health status into account when setting capitation rates', *Journal of the American Medical Association*, 276(16):1316-1321, October.

Frank, R.G., Glazer, J. and McGuire, T. G. (1998) 'Measuring adverse selection in managed health care', NBER working paper, No. 6825, December.

Frank, R.G., McGuire, T. G., Bae, J. P. and Rupp, A. (1997) 'Solutions for adverse selection in behavioral health care', *Health Care Financing Review*, Spring, 18(3):1-14.

Garber, A.M., MaCurdy, T.E. and McClellan, M.A. (1998) 'Persistence of Medicare expenditures among elderly beneficiaries', in A.M. Garber, ed., *Frontiers in Health Policy Research I*. Cambridge, MA: MIT Press, 153-180.

Gauthier, A.K., Lamphere, J.A. and Barrand, N.L. (1995) 'Risk selection in the health care maket: a workshop overview', *Inquiry*, 32:14-22.

Glazer, J. and McGuire, T.G. (forthcoming), 'Optimal risk adjustment in markets with adverse selection: an application to managed care', *American Economic Review*.

Greenwald, L.M., Esposito, A., Ingber, M.J. and Levy, J.M. (1998) 'Risk adjustment for the Medicare program: lessons learned from research and demonstrations', *Inquiry*, 35:193-209.

Gruenberg, L., Wallack, S.S. and Tompkins, C.P. (1986) 'Pricing strategies for capitated delivery system', *Health Care Financing Review*, Annual Supplement, 35-44.

Gruenberg, L., Tompkins, C. and Porell, F. (1989) 'The health status and utilization patterns of the elderly: implications for setting Medicare payments to HMO's', in R.M. Scheffler and L.F. Rossiter, eds., *Advances in health economics and health services research*, Vol. 10. JAI Press, Greenwich, 41-73.

Gruenberg, L. E., Kaganova, M. Hornbrook. (1996) 'Improving the AAPCC with health status measures from the MCBS', *Health Care Financing Review*, 17:59-76.

Ham, C. (1996) 'Managed markets in health care: the U.K. experiment', *Health Policy*, 35:279-292.

Hamilton, G.J. (ed) (1997) *Proceedings of the AIM-symposium 'Competition and solidarity*. published by the Alliance Nationale des Mutualités Chrétiennes (ANMC) of Belgium, Brussles, March.

Hay, J.W. and Olsen, R. J. (1984) 'Let them eat cake: a note on comparing alternative models of the demand for medical care', *Journal of Business and Economic Statistics*, 2:279-282.

Hellinger, F.J. (1995) 'Selection bias in HMOs and PPOs: a review of the evidence', *Inquiry* , 32: 135-142.

Hornbrook, M.C. and Goodman, M. J. (1995) 'Assessing relative health plan risk with the Rand-36 health survey', *Inquiry*, 32:56-74.

Hornbrook, M.C. and Goodman, M.J. (1996) 'Chronic disease, functional health status, and demographics: a multi-dimensional approach to risk adjustment', *Health Services Research*, 31(3):283-307.

Hornbrook, M.C., Goodman, M.J. and Bennett, M.D. (1991) 'Assessing health plan case mix in employed populations: ambulatory morbidity and prescribed drug models', in R.M. Scheffer and L.F. Rossiter, eds., *Advances in health economics and health services research*, Vol. 12. JAI Press, Greenwich, 197-232.

Hornbrook, M., Meenan, R., Bachman, D. et al. (1998) 'Forecasting health plan liability for defined populations: A Global Risk-Assessment Model', Working paper, Kaiser Permanente, June.

Iezzoni, L. ed. (1994) *Risk adjustment for measuring health care outcome*. Ann Arbor, Michigan: Health Administration Press.

Ingber, M.J. (1998) 'The current state of risk adjustment technology for capitation', Journal of Ambulatory Care Management, 21(4):1-28.

Jensen, G. and Morrisey, M. (1990) 'Group health insurance: a hedonic approach.' *Review of Economics and Statistics*, 72(1):38-44.

Kahn, J.G., Luft, H.and Smith, M.D. (1995) 'HIV risk adjustment: issues and proposed approaches', *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology*, 8(Suppl.1):S53-S66.

Keeler, E.B., Carter, G.M. and Trude, S. (1988) 'Insurance aspects of DRG outlier payments', *Journal of Health Economics*, 7:193-214.

Keeler, E.B., Carter, G.M. and Newhouse, J.P. (1998) 'A model of the impact of reimbursement schemes on health plan choice', *Journal of Health Economics*, 17:297-320.

Kennedy, A. (1996) 'Private health insurance in Ireland; the advent of a competitive market and a risk equalisation scheme', in J. Hermess, ed., *Risk structure compensation*. Proceedings of an AIM workshop, Maastricht.

Kesenne, J. (1996) 'Health care reform and risk structure compensation: the case of Belgium', in J. Hermesse, ed., *Risk structure compensation*. Proceedings of an AIM workshop, Maastricht.

Kronick, R., Zhou, Z. and Dreyfus, T. (1995) 'Making risk adjustment work for everyone', *Inquiry*, 32:41-55.

Kronick, R., Dreyfus, T., Lee, L. and Zhou, Z. (1996) 'Diagnostic risk adjustment for Medicaid: The disability payment system', *Health Care Financing Review*, 17:7-34.

Lamers, L.M. (1998a) 'Risk adjusted capitation payments: developing a diagnostic cost groups classification for the Dutch situation', *Health Policy*, 45:15-32.

Lamers, L.M. (1998b) 'Risk adjusted capitation based on the diagnostic cost group model: an empirical evaluation with health survey in formation', *Health Services Research*.

Lamers, L.M. (1998c) *Pharmacy costs groups: a risk adjuster for capitation payments based on the use of prescribed drugs?*. Report, Institute Health Policy and Management Erasmus University Rotterdam.

Lamers, L.M. (1999) 'The simultaneous predictive accuracy for future health care expenditures of DCGs, PCGs and prior costs', paper to be presented at the iHEA-Conference, 6-9 June 1999, Rotterdam.

Lamers, L.M. and van Vliet, R.C.J.A. (1996) 'Multiyear diagnostic information from prior hospitalizations as a risk adjuster for capitation payments', *Medical Care*, 34:549-561.

Lamphere, J.A., Neuman, P., Langwell, K. and Sherman, D. (1997) 'The surge in Medicare managed care: an update', *Health Affairs*, 16:127-133.

Lee, C. and Rogal, D. (1997) *Risk adjustment: a key to changing incentives in the health insurance market, a special report*. Robert Wood Johnson Foundation, Alpha Center, Washington, March.

Lichtenstein, R., Thomas, J.W., Adams-Watson, J., Lepkowski, J. and Simone, B. (1991) 'Selection bias in TEFRA at-risk HMOs', *Medical Care*, 29:318-331.

Light, D.W. (1992) 'The practice and ethics of risk-rated health insurance', *Journal of the American Medical Association*, 267:2503-2508.

Londono, J.L. (1996) 'Managed competition in the tropics?', paper presented at the International Health Economics Association Inaugural Conference, Vancouver, May.

Lubitz, J. (1987) 'Health status adjustments for Medicare capitation', *Inquiry*, 24:362-375.

Lubitz, J., Beebe, J. and Riley, G. (1985) 'Improving the Medicare HMO payment formula to deal with biased selection', in R.M. Scheffler and L.F. Rossiter, eds., *Advances in Health Economics and Health Services Research*, Vol. 6. JAI Press, Greenwich, 101-122.

Luft, H.S. (1982) 'Compensating for biased selection in health insurance', *Milbank Quarterly*, 64: 566-591.

Luft, H.S. (1986) 'Compensating for biased selection in health insurance', *Milbank Quarterly*, 64:566-591.

Luft, H.S. (1987) 'Health maintenance organizations and the rationing of medical care', *Milbank Memorial Fund Quarterly/ Health and Society*, 60:268-306.

Luft, H.S. (1996) 'Modifying managed competition to address cost and quality', *Health Affairs*, 15(1):23-38.

Luft, H.S. and Miller, R.H. (1988) 'Patient selection in a competitive health care system', *Health Affairs*, 7 (Summer):97-119.

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

Manning, W.G. and Marquis, M.S. (1996) 'Health insurance: the tradeoff between risk pooling and moral hazard', *Journal of Health Economics*, 15:609-639.

Manning, W.G. et al. (1987) 'Health Insurance and the demand for medical care: evidence from a randomized experiment', *American Economic Review*, 77:251-277.

Marquis, M.S. (1992) 'Adverse selection with multiple choice among health insurance plans: a simulation analysis', *Journal of Health Economics*, 11:129-151.

Matsaganis, M. and Glennerster, H. (1994) 'The threat of cream skimming in the post-reform NHS', *Journal of Health Economics*, 13:31-60.

McCall, N. and Wai, H.S. (1983) 'An analysis of the use of Medicare services by the continuously enrolled aged', *Medical Care*, 21:567-585.

McCarthy, T., Davies, K., Gaisford, J. and Hoffmeyer, U. (1995*) Risk adjustment and its implications for efficiency and equity in health care systems.* National Economic Research Associates (NERA), Los Angeles/ London, December.

McClure, W. (1984) 'On the research status of risk adjusted capitation rates', *Inquiry*, 21:205-213.

Medicare Payment Advisory Commission (MEDPAC) (1998) 'Report to Congress: Medicare Payment Policy'. March.

Morrisey, M. (1992) *Price Sensitivity in Health Care: Implications for Health Care Policy.* Monograph published by National Federation of Independent Businesses.

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

Newhouse, J.P. (1982) 'Is competition the answer?', *Journal of Health Economics*, 1:109-115.

Newhouse, J.P. (1984) 'Cream skimming asymmetric information, and a competitive insurance market', *Journal of Health Economics*, 3:97-100.

Newhouse, J.P. (1986) 'Rate adjusters for Medicare capitation', *Health Care Financing Review*, Annual Supplement , 45-55.

Newhouse, J.P. (1994) 'Patients at risk: health reform and risk adjustment', *Health Affairs*, 13:132-146.

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. (1998) 'Risk adjustment: where are we now?', *Inquiry*, 35:122-131.

Newhouse, J.P., Manning, W.G., Keeler, E.B. and Sloss, E.M. (1989) 'Adjusting capitation rates using objective health measurers and prior utilization', *Health Care Financing Review*, 10(3):41-54.

Newhouse, J.P., Sloss, E.M., Manning, W.G. and Keeler, E.B. (1993) 'Risk adjustment for a children's capitation rate', *Health Care Financing Review*, 15(1):39-54, Fall.

Newhouse, J.P., Buntin, M.B. and Chapman, J.D. (1997) 'Risk adjustment and Medicare: taking a closer look', *Health Affairs*, 16:26-43.

Pauly, M.V. (1980) *Doctors and their workshops: economic models of physician behavior*. University of Chicago Press, Chicago, IL.

Pauly, M.V. (1984) 'Is cream skimming a problem for the competitive medical market?', *Journal of Health Economics*, 3:87-95.

Pauly, M.V. (1986) 'Taxation, health insurance, and market failure', *Journal of Economic Literature*, 24(3):629-675.

Pauly, M.V. (1988) 'Efficiency, equity and costs in the U.S. health care system', in C.C. Havighurst et al., eds., *American health care: what are the lessons for Britain*?. London, IEA Health Unit, 23-45.

Pauly, M.V. (1992) 'Risk variation and fallback insurers in universal coverage insurance plans', *Inquiry*, 29:137-147.

Physician Payment Review Commission (1994) 'A comparison of alternative approaches to risk measurement'. Selected external research series, number 1, Washington D.C.

Pope, G.C., Adamache, K.W., Khandker, R.K. and Walsh, E.G. (1998a) 'Evaluating Alternative Risk Adjusters for Medicare', *Health Care Financing Review*.

Pope, G., Ellis, R., Liu, C.F. , Ash, A., Iezzoni, L., Ayanian, J., Bates, D. and Burstin, H. (1998b) *Revised Diagnostic Cost Group (DCG)/Hierarchical Coexisting Conditions (HCC) Models for Medicare Risk Adjustment.* Final Report to HCFA, Health Economics Research, Waltham, MA, February.

Pope, G.C., Liu, C.F., Ellis, R.P., Ash, A. S., Iezzoni, L., Ayanian, J. Z., Bates, D.W., Burstin, H., Adamache, K. and Gilman, B. (1999*) Principal Inpatient Diagnostic Cost Group Models for Medicare Risk Adjustment*. Final Report to HCFA, Health Economics Research, Waltham MA, February.

Porell, F.W., Turner, W.M. (1990) 'Biased selection under the senior health plan prior use capitation formula', *Inquiry*, 27(1):39-50.

Price, J.R., Mays, J.W. and Trapnell, G.R. (1983) 'Stability in the Federal Employees Health Benefits Program', *Journal of Health Economics*, 2:207-223.

Price, J.R. and Mays, J.W. (1985) 'Biased selection in the Federal Employees Health Benefits Program', *Inquiry*, 22:67-77.

Riley, G., Tudor, C., Chiang, Y. et al. (1996) 'Health status of Medicare enrollees in HMOs and fee-for-service in 1994', *Health Care Financing Review*, 16(4):65-76, Summer.

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.

Rogal, D.L., Gauthier, A.K. (1998) 'Are health-based payments a feasible tool for addressing risk segmentation?', *Inquiry*, 35:115-121.

Rothschild, M. and Stiglitz, J. (1976) 'Equilibrium in competitive insurance markets: An essay on the economics of imperfect information', *Quarterly Journal of Economics*, 90:629-649.

Schneider, B. (1996) 'Risk structure compensation in Switzerland', in J. Hermesse, ed., *Risk structure compensation*. Proceedings of an AIM workshop, Maastricht.

Schokkaert, E., Kestens, P., Dhaene, G., Lustig, H., van de Voorde, C., Laasman, J.M.,,de Lange, B. and Lona, M. (1996) 'Normuitkeringen voor de Belgische Ziekenfondsen: de eerste fase', *Openbare Uitgaven*, 28(4):195-204.

Schokkaert, E., Dhaene, G. and van de Voorde, C. (1998) 'Risk adjustment and the trade-off between efficiency and risk selection; an application of the theory of fair compensation', *Health Economics*, 7:465-480.

Schokkaert, E. and van de Voorde, C. (1998) 'Risk adjustment and the fear of markets: the case of Belgium', University of Leuven. Unpublished working paper.

Schut, F.T. (1995) 'Competition in the Dutch health care sector', unpublished doctoral dissertation, Erasmus University Rotterdam.

Selden, T.M. (1990) 'A model of capitation', *Journal of Health Economics*, 9:397-409.

Sheiman, I. (1994) 'Forming the system of health insurance in the Russian Federation', *Social Science and Medicine*, 39:1425-1432.

Sheiman, I. (1995) 'New methods of financing and managing health care in the Russian Federation', *Health Policy*, 32:167-180.

Shen, Y. and Ellis, R. P. (1998) 'Cost-Minimizing Risk Adjustment and Selection', Boston University working paper, October.

Sheldon, T.A., Smith, P., Borrowitz, M., Martin, S. and Carr Hill, R. (1994) 'Attempt at deriving a formula for setting general practitioner fundholding budgets', *British Medical Journal*, 309:1059-1064.

Shewry, S., Hunt, S., Ramey, J. and Bertko, J. (1996) 'Risk adjustment: the missing piece of market competition', *Health Affairs*, 15:171-181, Spring.

Shirom, A. (1995) 'The Israeli health care reform: a study of an evolutionary major change', *International Journal of Health Planning and Management*, 10:5-22.

Swartz, K. (1995) 'Reducing risk selection requires more than risk adjustments', *Inquiry*, 32:6-10.

Thomas, J.W., Lichtenstein, R., Wyszewianski, L. and Berki, S. (1983) 'Increasing Medicare enrollment in HMO's: The need for capitation rates adjusted for health status', *Inquiry*, 20:227-239.

Thomas, J.W. and Lichtenstein, R. (1986) 'Including health status in Medicare's adjusted average per capita cost capitation formula', *Medical Care*, 24:259-275.

Tolley, H.D. and Manton, K.G. (1984) 'Assessing health care costs in the elderly', *Transactions of the Society of Actuaries*, 36:579-603.

United States General Accounting Office (1994) 'Medicare: changes to HMO rate setting method are needed to reduce program costs', (GAO/HEHS-94-119) Washington DC, USA, (September).

United States General Accounting Office (1996) 'Private health insurance; millions relying on individual market face cost and coverage trade-offs', (GAO/HEHS-97-8) Washington DC, USA, (November).

United States General Accounting Office (1998) 'Health insurance standards; new federal law creates challenges for consumers, insurers, regulators', (GAO/HEHS-98-67) Washington DC, USA, (February).

Van Barneveld, E.M. (1999) 'Pooling as a supplement to capitation payments in a regulated competitive individual health insurance market', working paper (Ph.D. dissertation) in progress.

Van Barneveld, E.M., van Vliet, R.C.J.A. and van de Ven, W.P.M.M. (1996) 'Mandatory high-risk pooling: an approach to reducing incentives for cream skimming', *Inquiry*, 33:133-143.

Van Barneveld, E.M., van Vliet, R.C.J.A. and van de Ven, W.P.M.M. (1997) 'Risk adjusted capitation payments for catastrophic risks based on multi-year prior costs', *Health Policy*, 39:123-135.

Van Barneveld, E.M., Lamers, L.M., van Vliet, R.C.J.A. and van de Ven, W.P.M.M. (1998) 'Mandatory pooling as a supplement to risk adjusted capitation payments in a competitive health insurance market', *Social Science and Medicine*, 47:223-232.

Van de Ven, W.P.M.M. and Schut, F.T. (1994) 'Should catastrophic risks be included in a regulated competitive health insurance market?', *Social Science and Medicine*, 39:1459-1472.

Van de Ven, W.P.M.M. and van Vliet, R.C.J.A. (1992) 'How can we prevent cream skimming in a competitive health insurance market? The great challenge for the '90's', in P. Zweifel and H.E. French, eds., *Health Economics Worldwide*. Kluwer Academic Publishers, the Netherlands, 23-46.

Van de Ven, W.P.M.M. and Van Vliet, R.C.J.A. (1995) 'Consumer information surplus and adverse selection in competitive health insurance markets: an empirical study', *Journal of Health Economics*, 14:149-169.

Van de Ven, W.P.M.M., van Vliet, R.C.J.A., van Barneveld, E.M. and Lamers, L.M. (1994) 'Risk adjusted Capitation: Recent Experiences in The Netherlands', *Health Affairs*, 13: 118-136.

Van de Ven.,W.P.M.M., van Vliet, R.C.J.A., Schut, F.T. and van Barneveld, E.M. (1997) 'Access to coverage in a competitive individual health insurance market: via premium regulation or via risk adjusted vouchers?', report, September, Department of Health Policy and Management, Erasmus University Rotterdam.

Van Vliet, R.C.J.A. (1992) 'Predictability of individual health care expenditures', *The Journal of Risk and Insurance*, 59(3):443-460.

Van Vliet, R.C.J.A. (1997) 'Beperkte vormen van pooling als aanvulling op demografisch bepaalde normuitkeringen (Mandatory pooling as remedy for crude risk adjusted capitation payments to health plans)', *Tijdschrift Sociale Gezondheidszorg*, 75:244-251.

Van Vliet, R.C.J.A. (1999) 'A statistical analysis of mandatory pooling across health insurers', report, Department of Health Policy and Management, Erasmus University, Rotterdam.

Van Vliet, R.C.J.A. and Lamers, L.M. (1998) 'The high costs of death: should health plans get higher payments when members die?', *Medical Care* (forthcoming).

Van Vliet, R.C.J.A. and van de Ven, W.P.M.M. (1992) 'Towards a capitation formula for competing health insurers; an empirical analysis', *Social Science and Medicine*, 34:1035-1048.

Van Vliet, R.C.J.A. and van de Ven, W.P.M.M. (1993) 'Capitation payments based on prior hospitalization', *Health Economics*, 2:177-188.

Von Korff, M., Wagner, E.H. and Saunders, K. (1992) 'A chronic disease score from automated pharmacy data', *Journal of Clinical Epidemiology*, 45(2):197-203.

Ware, J.E. and Sherbourne, C.D. (1992) 'The MOS 36 item short form health survey (SF-36)', *Medical Care*, 30:473-483A.

Weiner, J.P., Dobson, A., Maxwell, S. et al. (1996) 'Risk adjusted Medicare capitation rates using ambulatory and inpatient diagnoses', *Health Care Financing Review*, 17:77-100.

Weiner, J.P., Starfield, B., Steinwachs, D. and Mumford, L. (1991) 'Development and application of a population oriented measure of ambulatory care case mix', *Medical Care*, 29:452-472.

Welch, W.P. (1985) 'Regression towards the mean in medical care costs: Implications for biased selection in HMOs', *Medical Care*, 23:1234-1241.

White House Task Force on Health Risk Pooling for small-group health insurance (1993) Washington D.C.

Wilson, C. (1977) 'A model of insurance markets with incomplete information', *Journal of Economic Theory*, 12:167-207.

Wouters, A.V. (1991) 'Disaggregated annual health services expenditures: their predictability and role as predictors', *Health Services Research*, 26:247-272.

Zeckhauser, R. (1970) Medical insurance: a case study of the tradeoff between risk spreading and appropriate incentives', *Journal of Economic Theory*, 2(1):10-26.

Zellner, B.B., Haugen, D.K. and Dowd, B. (1993) 'A study of Minnosota's high-risk health insurance pool', *Inquiry*, 30:170-179.

Table 1. Effects of selection

Effects of *adverse selection:*

- high premiums for high-risk individuals;
- dependent upon the level of the contracting costs either the low-risk individuals or the high-risk individuals cannot obtain as much health plan coverage as they wish;
- welfare loss in the case of an unstable market (including bankruptcy of adversely selected health plans).

Effects of *cream skimming*:

- disincentive for the health plans to respond to the preferences of high-risk consumers;
- incentive to provide poor quality of care and poor service to high-risk individuals;
- disincentives for providers and health plans to acquire the best reputation for treating chronic illness;
- dependent upon the form of premium regulation (per health plan or nation-wide): high premiums for high-risk patients or bankruptcy of non-skimming selected health plans;
- investments in cream skimming have higher returns than investments in improving efficiency;
- investments in cream skimming (e.g., resources to identify and attract high-risk consumers) are a welfare loss.

Table 2
Health Plan Response to Incentives Created by the Way that Health Plans are Reimbursed

Choice of plan benefit features
- Deductibles or copayments for selected conditions.
- Coverage limits (lifetime or annual)
- Coverage of pharmaceuticals or other specific services
- Exclusions for preexisting conditions

Responses to regulated rate classes
- Efforts to attract more profitable rate classes such as:
  - family or individual contracts
  - employee or retiree
  - specific geographic area
- Selection of relative premiums by rate classes

Plan level efforts to attract profitable/avoid unprofitable enrollees
- Denying coverage ("medical underwriting")
- Canceling coverage
- Selective advertising
- Pre-enrollment screening
- Selective enrollment and disenrollment counseling

Changes in service offerings
- Selection of specialists to include or exclude from plan network
- Over-provision of services that attract profitable enrollees
- Underprovision of services that attract unprofitable enrollees
- Change of place of service to increase payments
- Unnecessary provision of services to code a diagnosis
- Change in timing of services to increase payment

Changes in diagnostic coding or other claims information
- Upcoding of diagnoses to more serious conditions
- Proliferation of diagnoses
- Fraudulent diagnostic coding
- Coding of "rule out" diagnoses

Attempts to influence survey-based health measures
- Enrollee coaching
- Nonrandom enrollee sampling
- Biased corrections for nonresponse

Table 3

Comparison of R^2 from various risk adjustment models from six papers

Study	Newhouse et al., 1989	Van Vliet and van de Ven, 1992	Fowles, Weiner, et al 1996[b]	Physician Payment Review Commission 1994	Pope et al, 1998a	Lamers, 1998b
Sample population	US Privately Insured	Netherlands	US HMO enrollees	US, Medicare	US Medicare	Netherlands sickness fund
Sample period	1974-1979	1981-82	1991-1993	1991-1992	1991-1993	1991-1994
Sample size	N=7,690	N = 20,000	N = 5780		N = 10,893	N=10,570
Age/Sex	0.016	0.028	0.058	0.016	0.007	0.038
All socioeconomic[a]		0.037				
Functional status[a]					0.0252	
Self reported chronic conditions[a]		0.071	0.111	0.032	0.0274	
Self reported health[a]	0.028			0.03	0.0311	
Short-Form 36 like[a]			0.111	0.033	0.0405	
Prior year spending[a]	0.064				0.0413	
Comprehensive survey[a]		0.114		0.062	0.0418	0.060
Diagnosis based[a]	0.045		.124[c]		0.0727[d]	0.080[e]
All variables[a]	0.09			0.07	0.0785	0.086

Notes:

[a] all models include age and sex as well as variables shown

[b] Dependent variable was truncated at $25,000, which inflates R^2

[c] ACG/ADG model

[d] DCG/HCC model

[e] Three-year DCG-model

Table 4. Forms of risk-sharing

	Reimbursement rate	Threshold	Ex-ante percentage of members to whom the risk-sharing applies
Proportional risk-sharing	r	0	100
Outlier risk-sharing	r	T	100
Risk-sharing for high-risks	r	T	p

Table 5. The practice of risk-adjustment and risk-sharing in 10 countries*

	Belgium	**Colombia**	**Czech Republic**	**Germany**	**Ireland**	**Israel**	**Netherlands**	**Russia**	**Switzerland**	**United Kingdom**
Risk-adjusters	age/gender region disability unemployment mortality	age/gender region	age	age/gender disability	age/gender hospitalization both weighed with current expenses	age	age/gender region disability	many different regional experiments	age/gender region	age/gender prior utilization local factors
Restrictions on premium contribution	community rating	zero premium contribution	community rating	community Rating	community rating	zero premium contribution	community rating	zero premium contribution	community rating per region	zero premium contribution
Risk-sharing	proportional risk-sharing, at least 85%	no	no	no	see risk-adjusters (above)	severe diseases (6 percent of expenses)	outlier risk-sharing & Proportional Risk-sharing	many different regional experiments	no	outlier risk-sharing (£6000) (mid 1990s)
Number of health plans	6	24	26	1200	2 (until 1997:1)	4	25	100s	166	2500 (early 1996)
Modality A or B	A	B	B	B	B	A	A	A	B	A
Open entry for new health plans? (subject to certain conditions)	no	yes	yes	yes	yes	yes	yes	yes	yes	yes
Open enrollment every month/…/year	quarter	year	year	year	year	half year	year	year	half year	no open enrollment guarantee
Is long-term care included in benefits package?`	yes	no	no	no	no	no	no	no	no	no
Mandatory or voluntary membership?	M	V	M	M	V	M	M	M	M	V
Year of implementation	1995	1994	1993	1994	1996	1995	1991	1993	1993	1991

* Source: Chernichovsky and Chinitz (1995), Files and Murray (1995), Ham (1996), Kennedy (1996), Kesenne (1996), Londono (1996), Matsaganis and Glennester (1994), McCarthy et al. (1996), Schneider (1995), Schokkaert et al. (1996), Sheiman (1994, 1995), Sheldon et al. (1994) and Shirom (1995).

Table 6. The practice of risk-adjustment and risk-sharing in US*

	Medicare program, HMOs in 1997	Medicare, proposed for HMOs in year 2000	Federal Employees Health Benefits' Programme (FEHBP)	New York State	Health Insurance Plan of California (HIPC)	Minnesota Buyers Health Care Action Group	Washington Health Care Authority
Risk-adjusters	age/gender region (county) institutional status welfare status	age/gender region (county) welfare status Principal Inpatient Diagnostic Cost Groups (PIPDCGs)	No risk-adjusters each consumer's subsidy is based on 60% of the average premium of the six largest plans	age/gender region	Gender number of children 120 marker diagnoses risk adjustment only applied if plan scores deviate from 1 by ±5.percent	ACGs	Age, Sex, employee status since 1989, DCGs announced for 2000
Restriction on premium contribution	community rating	community rating	community rating	community rating	premium contribution depends on age, region and family/single within a rate band (± 10%)	premium contributions set by competitive bidding	premium contributions based on competitive bids
Risk-sharing	no	no	no	condition-specific risk-sharing	no	Stoploss for catastrophic individuals	yes
Number of health plans	100s	100s?	100s	?	28	15	3
Modality A or B	A	A	A	B	B	A	A
Open entry for new health plans?	yes	yes	yes	yes	yes	yes	no
Open enrollment every month/…/year	month	month, with proposed transition to year	year	?	year	year	year
Is long term care included in benefits package	no	no	no	no	no	no	no
Mandatory or Voluntary membership?	V	V	V	V	V	V	V
Year of implementation	1972	2000	1960	1993	1992	1997	1989

* Source: Buchmueller (1997), Butler and Moffit (1995), Dunn (1998), Lee and Rogal (1997), McCarthy et al. (1995) and Shewry et al. (1996).

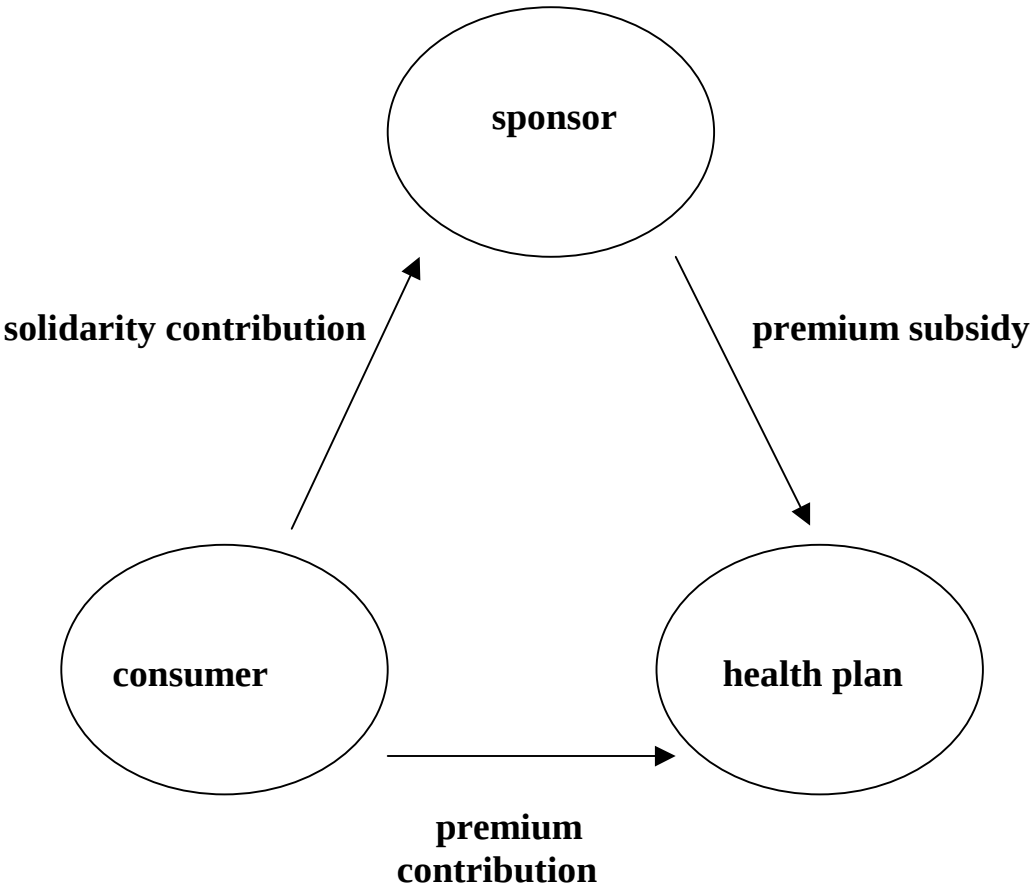


Figure 1. Risk adjustment system (Modality A)

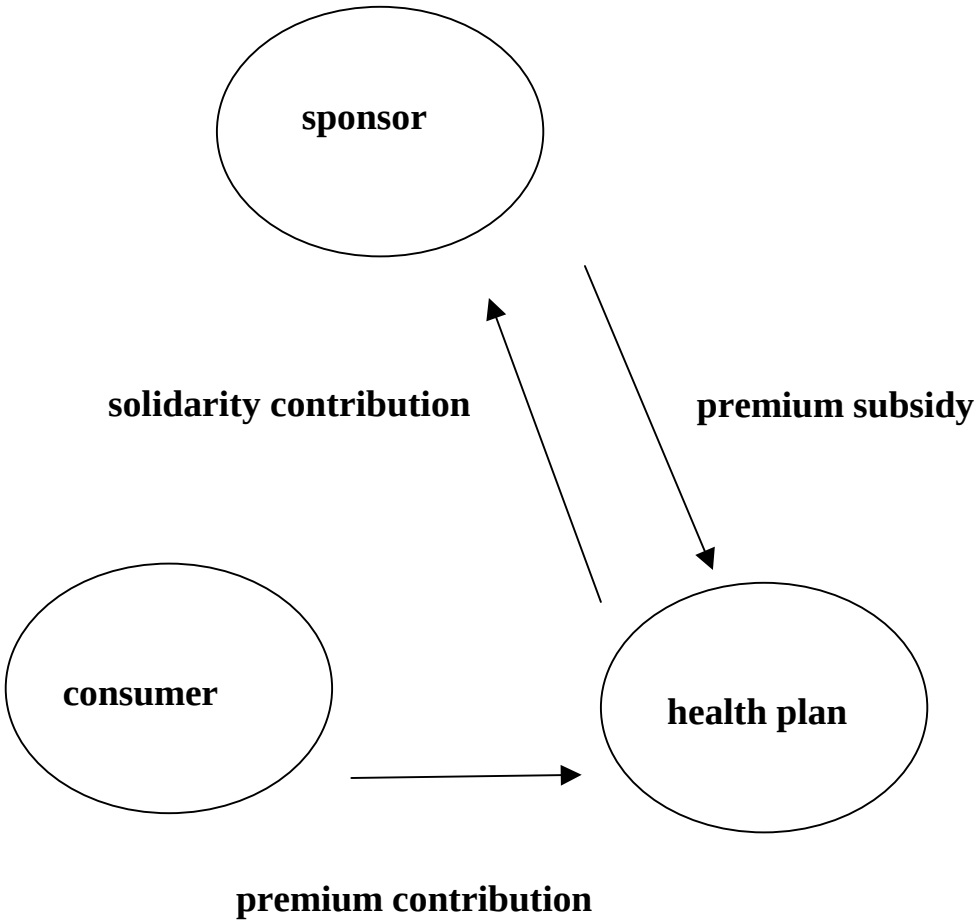


Figure 2. Risk adjustment system (Modality B)

Figure 3

Factors Explaining Variation in Health Spending

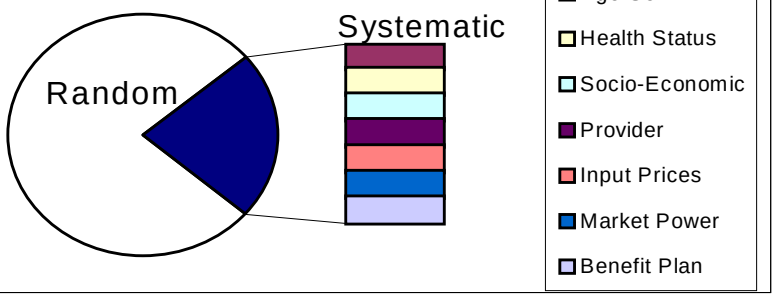


Figure 4A

Health Spending by Gender and Age in the USA

1.0 Million Privately Insured Individuals, 1992-93

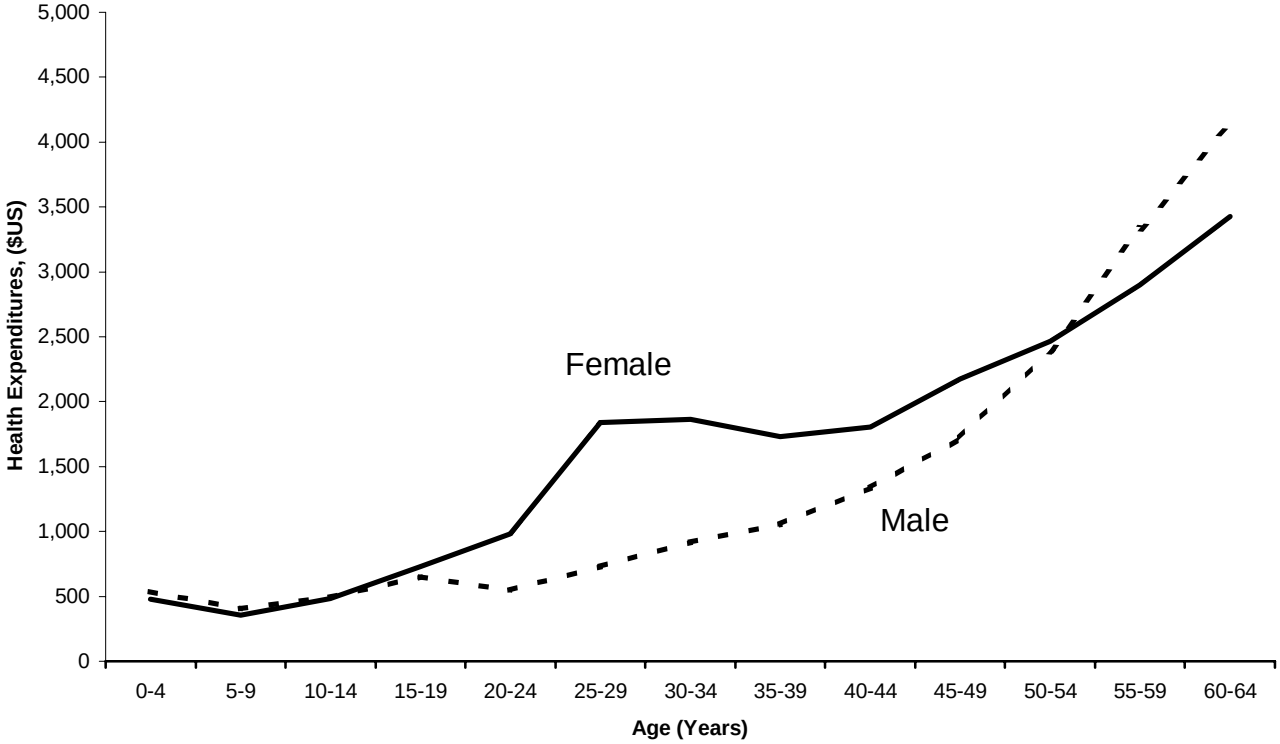


Figure 4B

Health Spending by Gender and Age in the USA

1.3 Million Medicaid Eligibles Under age 65, Michigan, 1991-92

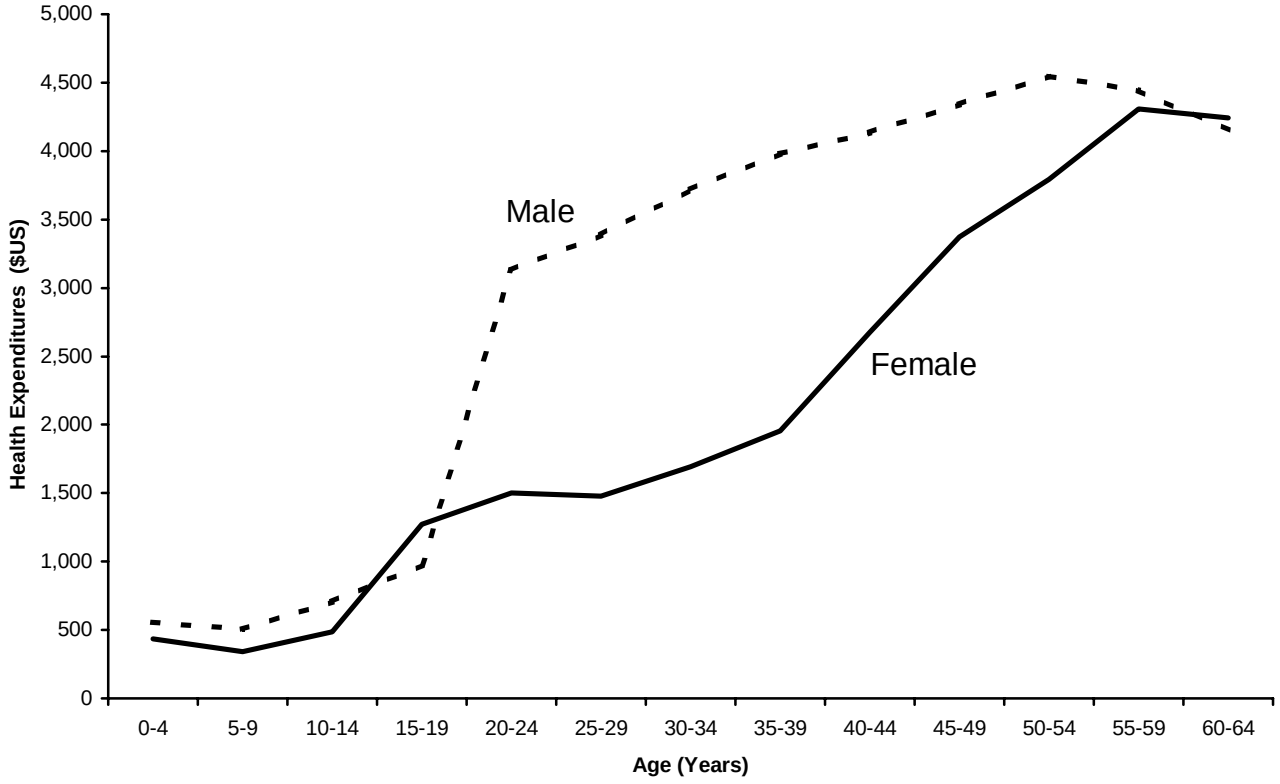


Figure 4C

Health Spending by Gender and Age in the Netherlands

9.6 Million Enrollees in Sickness Fund Basic Benefits Package, 1995.

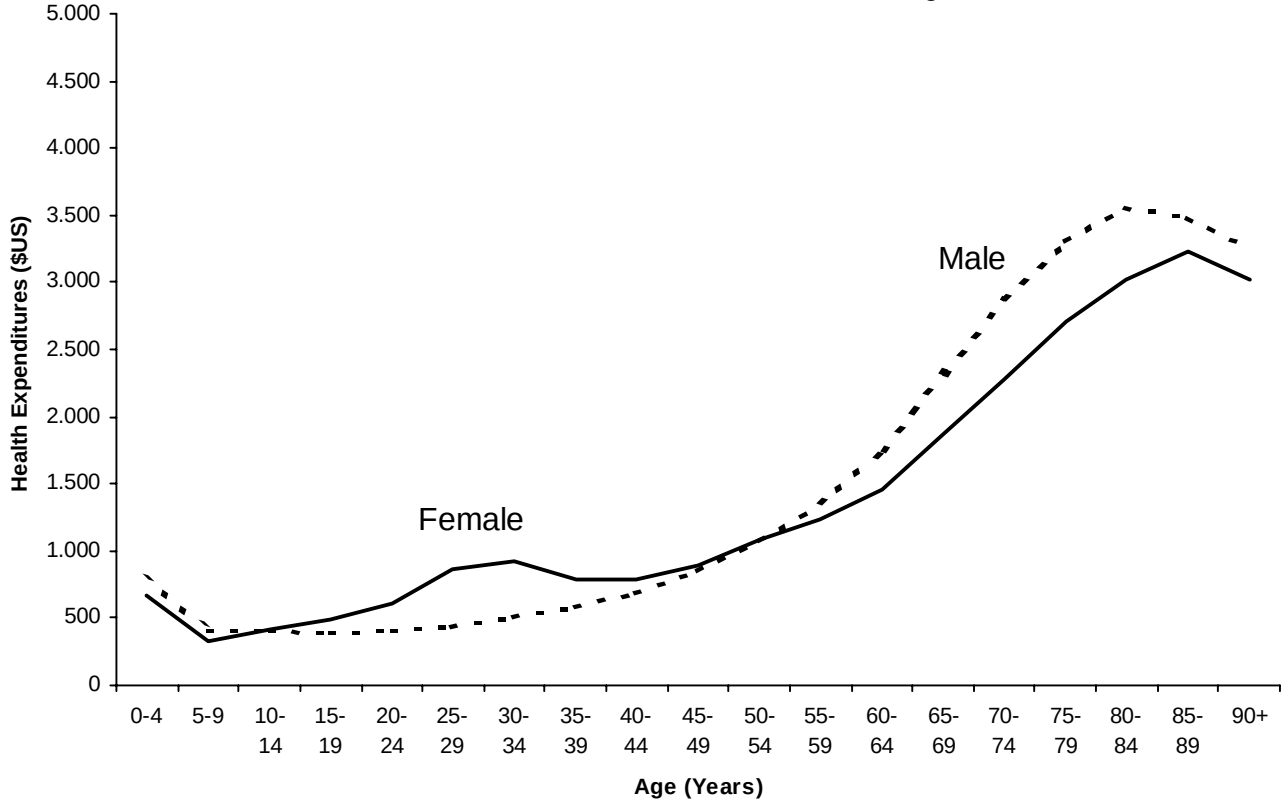
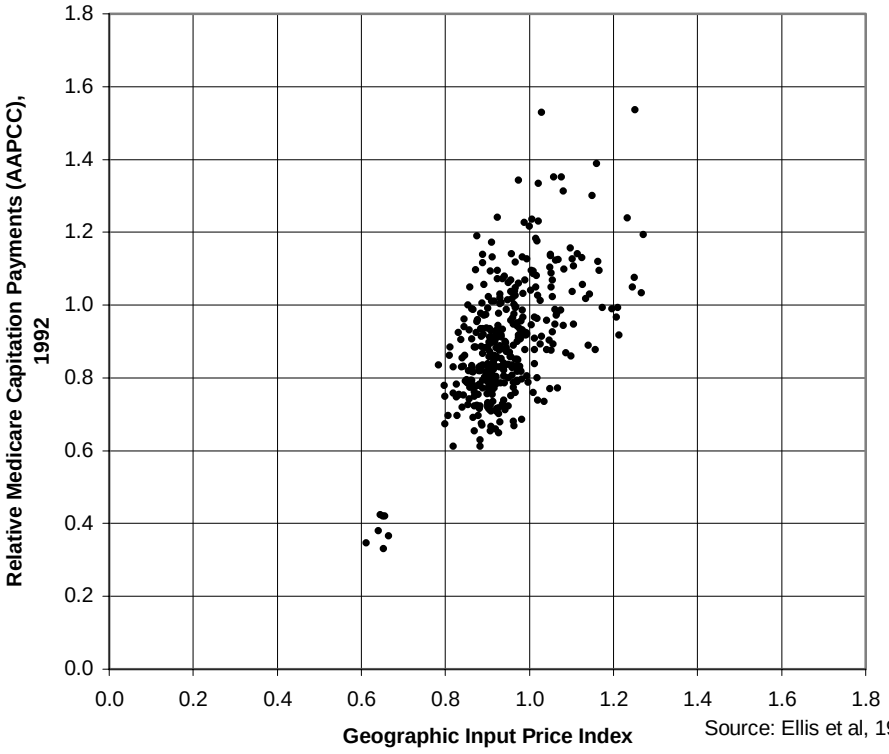


Figure 5

Scatter Plot of Geographic Input Price Index and Relative US Medicare Capitation Payment (AAPCC), 1992, (rho = .61)



Source: Ellis et al, 1996, Table G-1

Figure 6

Effects of Truncating Expenditures on R-Square

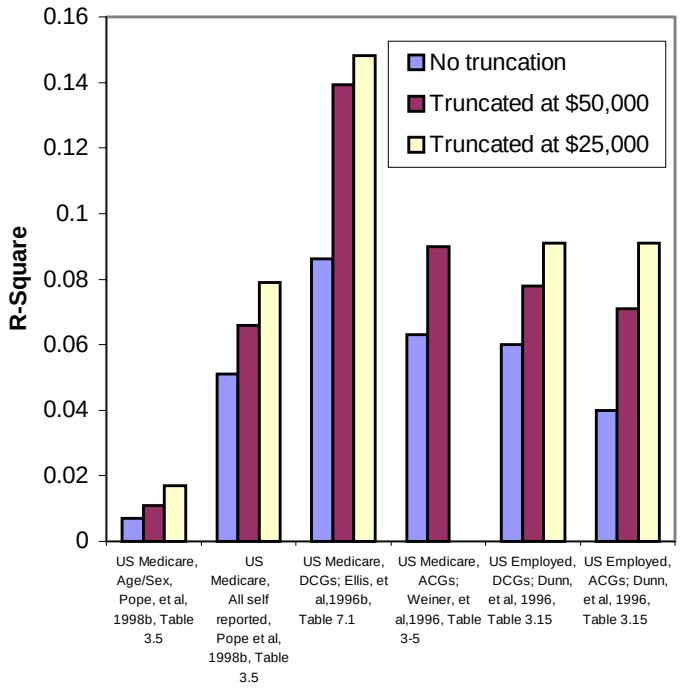


Figure 7

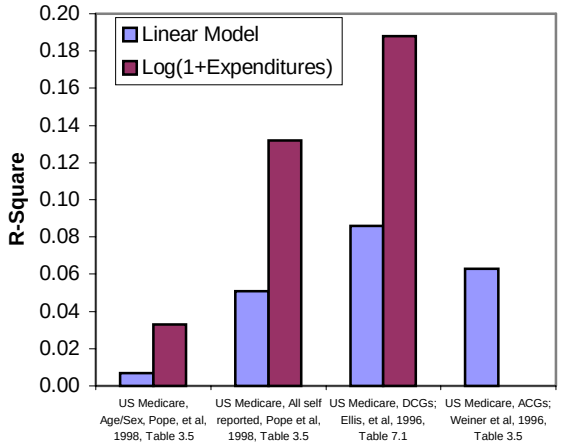
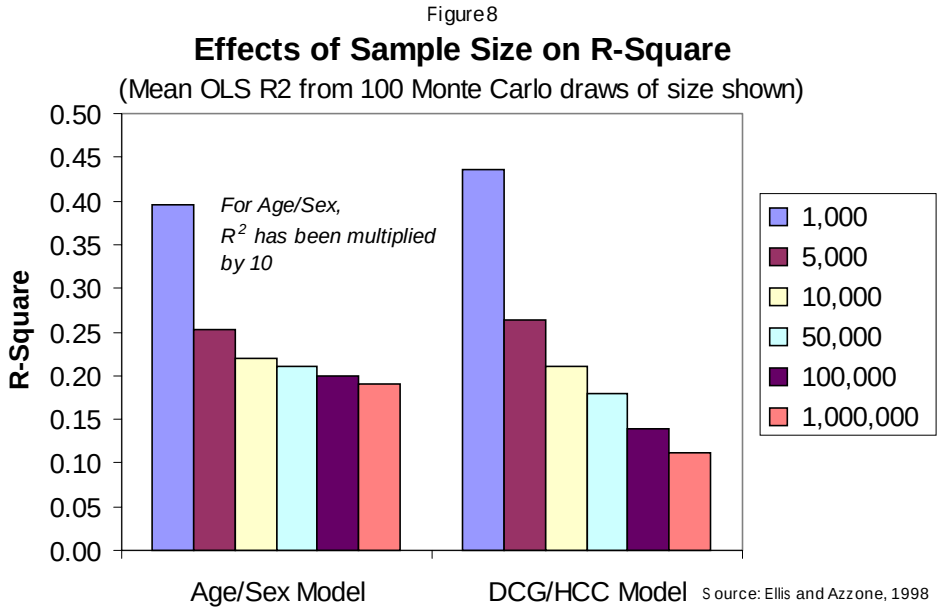

Effects of Log Transformation on R-Square
R-Square
0.20
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
Linear Model
Log(1+Expenditures)
US Medicare, Age/Sex, Pope, et al, 1998, Table 3.5
US Medicare, All self reported, Pope et al, 1998, Table 3.5
US Medicare, DCGs; Ellis, et al, 1996, Table 7.1
US Medicare, ACGs; Weiner et al, 1996, Table 3.5

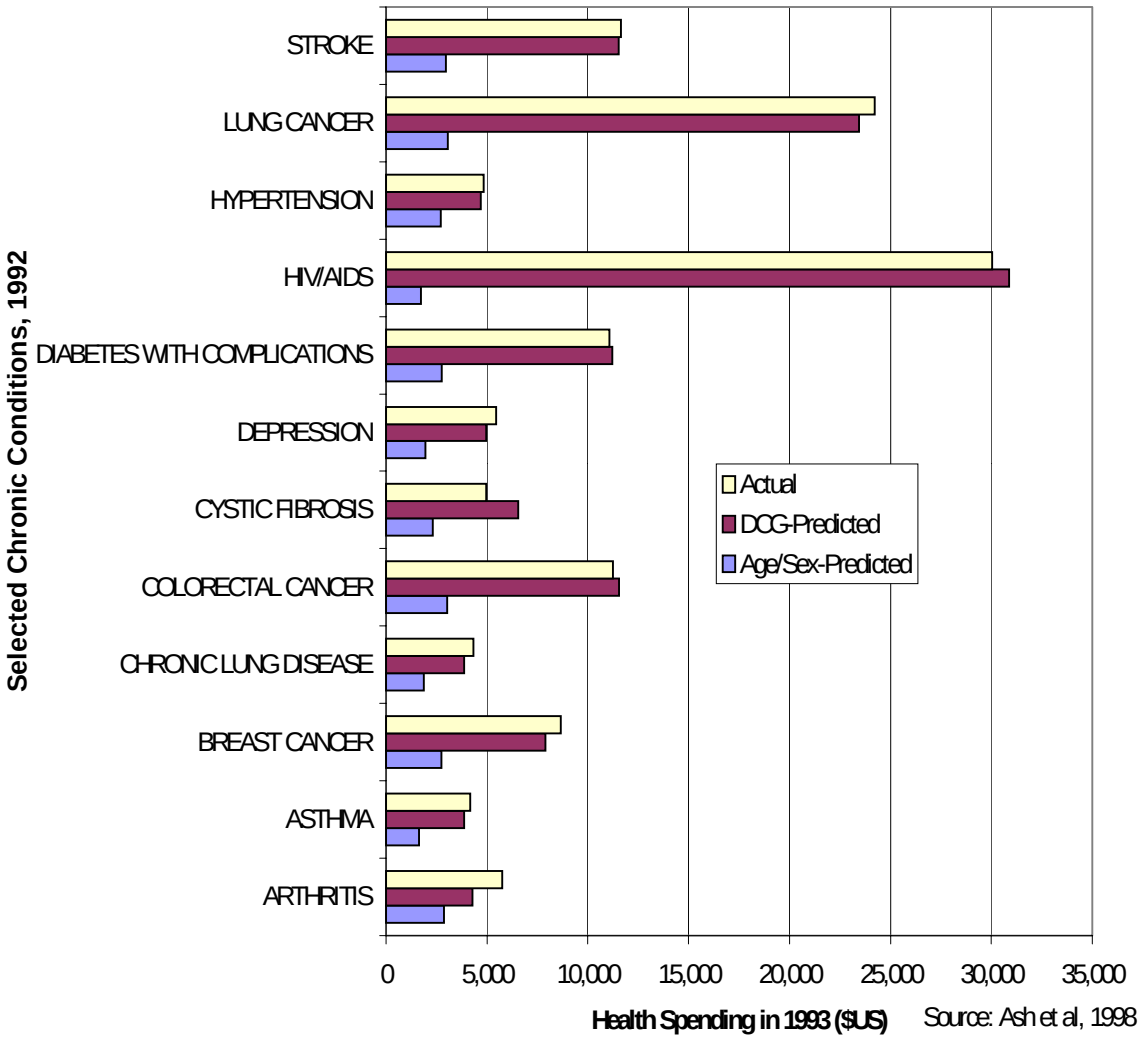
Figure 8



Source: Ellis and Azzone, 1998

Figure 9

Comparison of Actual Versus Predicted Health Spending
By Selected Chronic Conditions
US Privately Insured Sample (N=346,466)

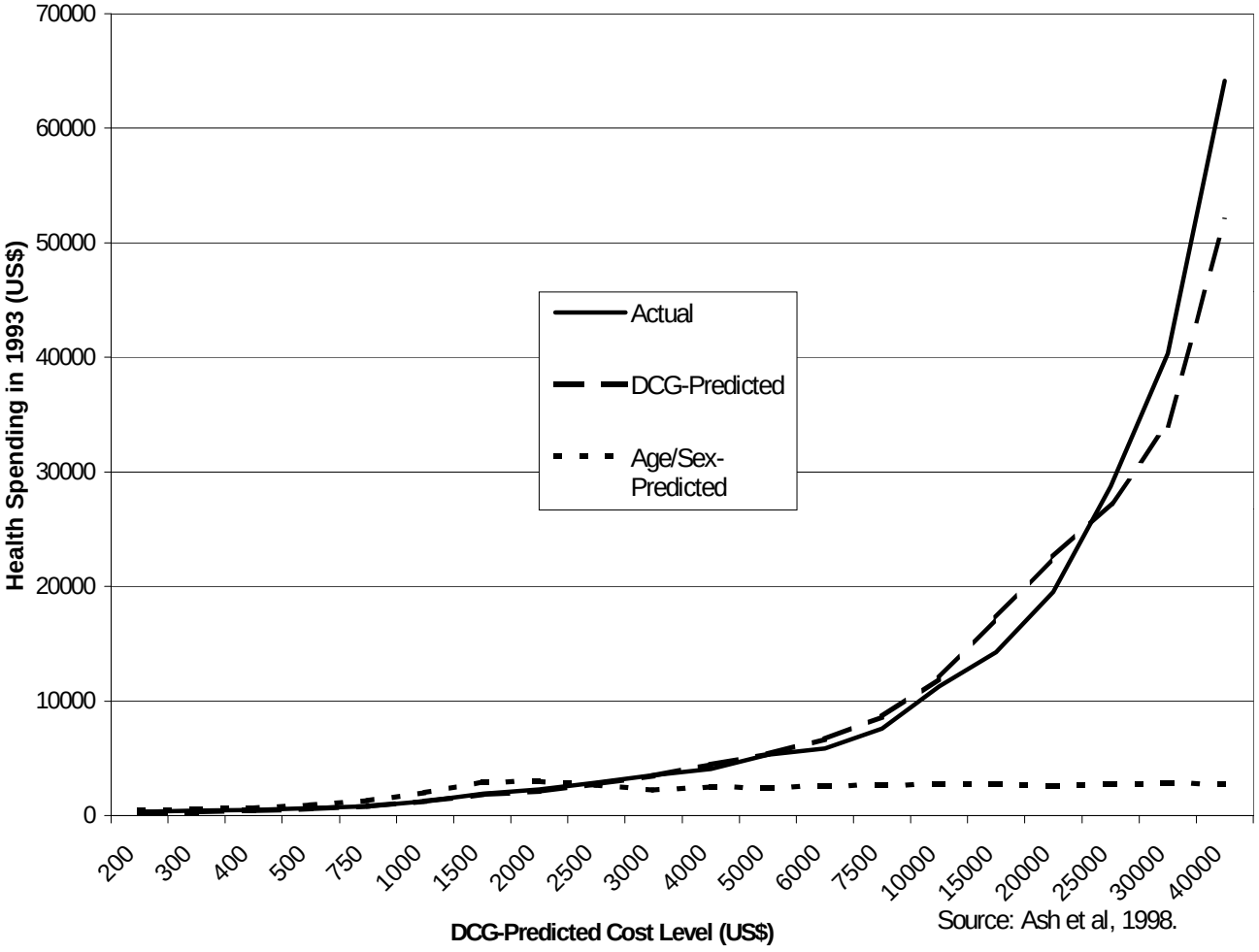


Source: Ash et al, 1998

Figure 10

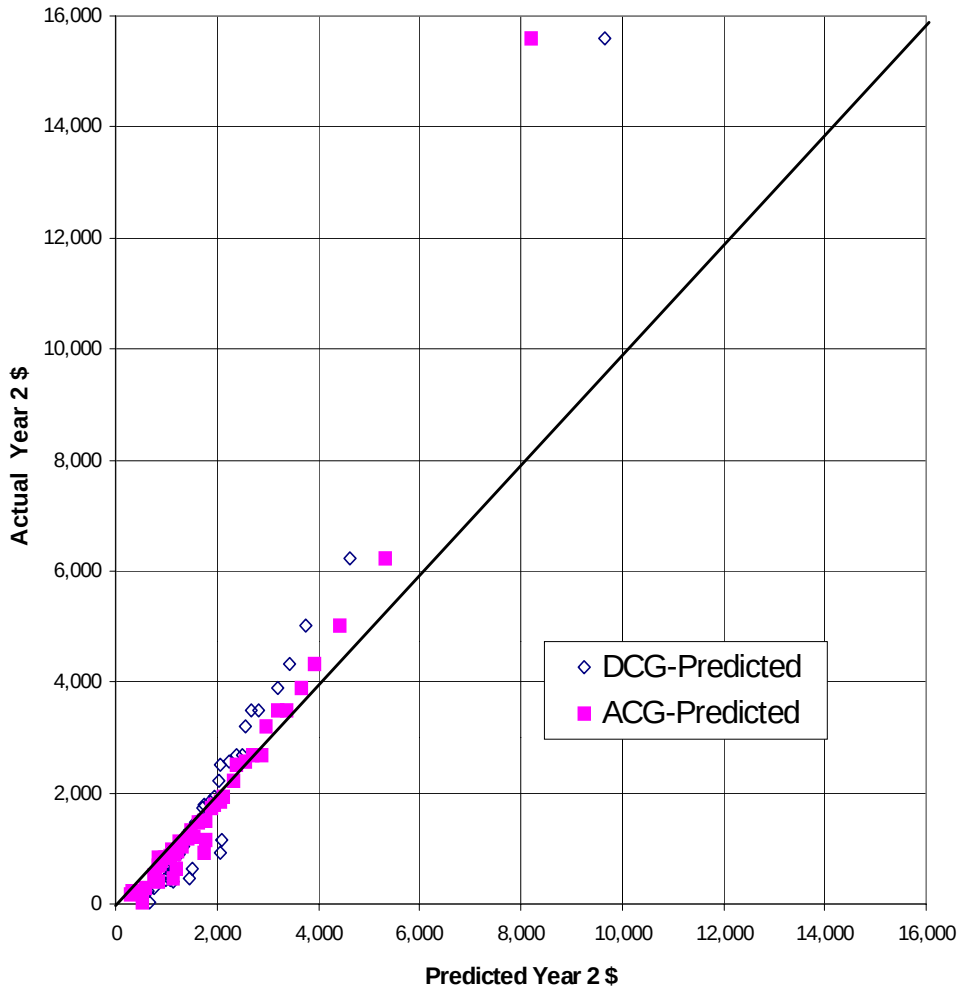
Comparison of Actual Versus Predicted Health Spending by DCG Predicted Cost Intervals

US Private Insured Sample (N=346,466)



Source: Ash et al, 1998.

Figure 11. Predicted Versus Actual Year 2 Costs With Each Observation Calculated for a 2%-ile Group Based on Year 1 Cost



Source: Ash and Byrne-Logan, 199