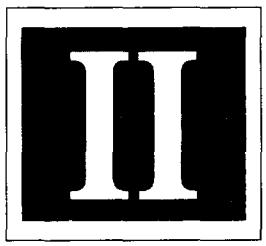

Background

In evaluating health risk assessment methods it is important to consider both the previous research in this area and how the methods have been applied in risk adjustment. We begin with a description of what, in practice, is being done.

A. Applications of Risk Adjustment

Actuaries have been using risk assessment for years in the pricing of health insurance using techniques such as age/sex rating, experience rating, and tier rating. HMO premiums for Medicare beneficiaries have also been risk-adjusted for more than a decade using age, sex, geography, welfare and institutional status, and so on, as described by the adjusted average per capita cost (AAPCC). In more recent years, alternative methods of risk assessment have been researched and developed, including models based on health status, as measured by service use and patient diagnoses. The federal government is currently exploring the use of health status measures as alternatives to the AAPCC.

Under the umbrella of health care reform, several states have either begun risk adjustment or are in the process of implementing risk adjustment legislation. As usual, with 51 different jurisdictions, many different approaches are being followed. The list below provides a brief description of the "natural experiments" now underway in several of these states:

1. New York

Beginning in April 1993, risk adjustment was imposed marketwide on all insurers selling policies in the individual and small group market in New York state. Because the new law required community rating without use of age as a premium rating factor, regulations were issued to assess the demographic distribution of insured enrollees (that is, age, gender, and family size); risk adjustment payments were then either paid (by insurers with lower risk populations) or received (by insurers with higher risk populations). The movement of risk payments for demographic reasons is made quarterly.

A second pool of funds is collected in New York through an assessment on premiums and distributed to insurers for a very limited list of high-cost-diagnosis cases. This system is intended to reimburse insurers that have enrolled individuals with very serious diseases or who have costly procedures (for example, several types of transplants), generally those which cost over $100,000 per case. The reimbursement level is set below average case cost to provide an incentive to insurers to closely manage the cases that are found.

Because of New York's strict community rating law, the demographic risk adjustment was needed to minimize the impact of community rating on those insurers that started with or attracted a higher risk (generally older) population. The high-cost-diagnosis pool was an additional attempt to pay insurers adequately when they were dealt a high-risk individual; although this is a retrospective payment method, it differs significantly from reinsurance since payment is linked to incidence of a case, not to funds spent on claims paid.

2. California

The health care reform law in California that established the small group purchasing pool (the Health Insurance Plan for California, or the HIPC) also provided authority for risk adjustment inside the HIPC but did not require it. Under the new legislation, insurers have a requirement to "guarantee issue" coverage but are allowed to continue using age, geography, and family size as rating factors within a +/−20% rating band

(which will be later reduced to +/−10%). The HIPC, however, has chosen to use only age, geography and family size without using the possible rating band to reflect prior claims experience or other factors.

Since the state can only require risk adjustment within the HIPC, the risk adjustment method chosen had to be administratively simple to avoid undue expense for HIPC insurers and had to be perceived as effective in adjusting between higher risk and lower risk insurers. While age adjustment was not necessary (unlike in New York, California insurers could age-rate their premiums), adjustments for gender-mix differences and family size differences are calculated using demographic information on HIPC enrollees. Because age-rating accounts for most of the difference in demographic characteristics, the amount of risk differences attributable to the other demographic factors is likely to be small.

The major risk assessment and adjustment element in the HIPC methodology involves the identification of individual enrollees who have been hospitalized in the prior year with a "marker diagnosis." HIPC consultants, working together with a committee from the insurers, identified approximately 50 diagnoses with high costs, all requiring an inpatient admission. Certain categories of high-cost cases, for example, traumas (which generally occur randomly) and mental illness (which had extremely wide differences in payment and treatment) were excluded. Each individual with one of these "marker diagnoses" was then assigned a weight (based on average costs derived from California health care experience in managed care plans); all other enrollees received an "average" weight. An individual HIPC insurer's average weight for all enrollees then determined a risk assessment score that was used to calculate proposed risk adjustment amounts (when combined with the demographic factors mentioned above).

For the 1995/96 HIPC contract year, a grant from the Robert Wood Johnson Foundation is funding a simulation of the risk adjustment process. After gathering data, the HIPC staff will inform insurers what the risk adjustment amounts would have been if the process had been operational. For the 1996/97 contract year, the HIPC is planning to implement risk adjustment, if risk assessment indicates that some insurers have risk outside a threshold of +/−5%. (If the decision were being made using 1995 information, risk adjustment would be triggered.)

Separately, the California Public Employees' Retirement System (CalPERS), California's pension and health care agency for state and local governmental employees, has been making use of demographic risk factors for several years to establish targets for rate negotiations. The success of CalPERS in negotiating several years of rate reductions may be due in part to its administrators' knowledge of which rates were justifiable and which appeared to be excessive.

The Pacific Business Group on Health (formerly the Bay Area Business Group on Health) also has a risk adjustment project in progress. This project is looking at prior claims methods, demographic methods, and self-reported health status (using versions of the SF-36 health status questionnaire) in the context of negotiations with insurers and HMOs on behalf of very large California employers.

3. Colorado

While Colorado has enacted legislation as part of two incremental health care reform bills that allows risk adjustment, there is relatively little ability for the state to pursue risk adjustment, which involves transferring funds from low-risk insurers to high-risk insurers. At the moment, insurers who are able to prove to the Division of Insurance that they will be financially impaired by the guarantee issue law may obtain relief from that requirement. There is, however, no apparent ability by the state to perform risk adjustment on insurers who are "low-risk winners" in the current environment.

4. Washington State

While much of Washington State's comprehensive health care reform was repealed earlier in 1995, the Health Care Authority (HCA) continues to study risk adjustment with the state's own employee group of nearly 300,000 individuals, which is administered by the HCA. A team from the University of Washington has just completed Phase I during which statistical risk assessment models were set up using data from six health plans (which enroll nearly 86% of the group) providing coverage to state employees and their dependents. Phase II (recently funded by the Robert Wood Johnson Foundation) will continue the work by designing and implementing a diagnostic risk-adjusted payment based on the Phase I models. An interesting part of this project is the assessment of a prescription drug usage method in addition to use of demographic

and more traditional diagnosis-based models (for example, ACGs, DCGs, etc.).

5. *Florida*

Florida has recently contracted with a major research firm to review the types of risk adjustment that may be possible under its reform legislation. Since both individuals and small groups are covered by guarantee issue requirements, results of the research should prove to be interesting.

6. *Kentucky*

In Kentucky, all individuals, small groups of less than 100 employees, and state employees are subject to risk adjustment if covered by an insured plan, whether in the alliance or outside. Age-rating is permitted in Kentucky. The risk adjustment mechanism is similar to New York's high-cost condition pool and demographic adjustments for COBRA, retirees, and gender are included.

In summary, previous applications of risk assessment and risk adjustment have involved a range of approaches. Efforts by states have typically employed demographic factors such as age, gender, family size, and geography, with some method of reinsurance or retrospective adjustment for high-cost cases. There have been a few instances where health status measures, such as diagnoses and conditions, have been employed, in particular, the marker diagnosis approach developed for the California HIPC. However, as of this time, no state has implemented complete diagnosis-based approaches such as ACGs or DCGs as part of a risk adjustment process. The practical applications of the diagnosis-based approaches have been restricted to setting capitation rates, provider profiling, and health services research.[1]

B. Prospective Versus Retrospective Applications of Risk Assessment and Risk Adjustment

Risk assessment can be performed prospectively or retrospectively, and the risk adjustment process can also be performed prospectively or retrospectively. Generally, prospective risk assessment uses the experience of one year, such as 1995, to predict the risk assessment attributes of an upcoming year, such as 1996. Prospective risk adjustment would apply the transfer payments to the insurance funding (or set payment rates for capitation purposes) in a prospective fashion—most typically, each carrier would build the expected risk adjustment transfer amounts into their premium rates. A true prospective methodology would imply that once the prospective assessments are used to determine transfers, there will be no further settlement.

Retrospective risk assessment would use the experience of one year to set the risk assessment attributes of that same year. Likewise, retrospective risk adjustment for a year would transfer payments between carriers based on actual usage and risk measured for that year. A retrospective settlement is an example of retrospective risk adjustment. A reinsurance system for large amount claims could also be considered retrospective.

A prospective methodology could be combined with a retrospective settlement. In this way, prospective methods would be used to determine premiums and initial risk transfers for the upcoming year. At the end of the year, retrospective adjustments could be made to reflect the risk of the individuals enrolled.

We explore both prospective and retrospective risk assessment in this study. These two approaches differ in the accuracy with which they measure differences in risk. They also differ in the incentives they provide to plans for efficiency and quality in medical care.

We discuss each of these issues later in this report.

C. Review of Previous Research on Risk Assessment

Our review of the literature on risk assessment addresses three interrelated questions of particular relevance to this project:

1. What methods have been proposed for risk assessment?
2. What criteria should be used to evaluate these methods?
3. How do the different methods compare?

We have not attempted to be exhaustive, but only to give the reader an overview of the literature sufficient to put our results in context and indicate their significance.

For those interested, a bibliography on risk assessment and risk adjustment is included in Appendix A of this report.

1. Alternative Methods Proposed for Risk Assessment

The objective of risk assessment is to measure the expected costs of subgroups in a population so they can be used to assign levels of risk to individuals or a group of enrollees. Existing models use two types of data as predictors of expected costs: demographic variables and health status. Demographic variables proposed include age, sex, family status, location, and welfare status, while measures of health status can range from self-reported health to diagnoses and prior use. Models incorporating health status also usually include demographic variables as predictors of costs.

The risk assessment methods that have been discussed in the literature include the AAPCC method that the Health Care Financing Administration (HCFA) uses to determine capitation rates for Medicare enrollees, ACGs, DCGs, payment amounts for capitated systems (PACS), self-reported health status measures, physiologic health measures, mortality patterns, prior use, the Robinson-Luft Multiequation Model, the New York State retrospective conditions/procedures payment method, and the more elaborate method using marker diagnoses developed in California. The New York State and California methods were described previously in our discussion of current state efforts on risk adjustment and will be further discussed in Chapter IV in the context of their methods for reinsurance and risk assessment for high-cost conditions.

Note also that the health insurance industry has for years (since the late 1960s) been doing risk assessment, particularly for individual major medical insurance. This process involves evaluating blood tests, analyzing attending physician statements, asking a series of medical history questions, and then using established guidelines that determine whether a person is 25% higher cost risk, 50% higher cost risk, and so on. One risk assessment system could involve an independent national body of health underwriters through which all applications could be required to pass in order to evaluate a relative risk.

We summarize briefly the remaining models below.

AAPCC. The Tax Equity and Fiscal Responsibility Act of 1982 (TEFRA) authorizes prospective per capita payment to HMOs and comprehensive medical plans (CMPs) for Medicare enrollees at a rate equal to 95% of the AAPCC. The AAPCC is based on average Medicare payments per fee-for-service beneficiary in a county, adjusted for differences in input prices. The amount thus determined is then adjusted according to the age, sex, welfare status, institutional status, and basis for Medicare eligibility (age, disability, or end-stage renal disease) of the beneficiary. Concern over its inadequacy, within HCFA as well as outside, has fostered much of the research on alternative risk adjusters (Newhouse et al., 1989).[2]

ACGs. ACGs were developed by a research team at Johns Hopkins University under the leadership of Jonathan Weiner (Starfield, Weiner, et al., 1991). Unlike many other risk assessment methods, ACGs were not developed as an alternative to the AAPCC, but initially as a method for predicting ambulatory care costs alone. However, since that time, they have been used for predicting both ambulatory and total medical expenditures.

ACGs are a diagnosis-based measure, where individuals are assigned to a group based on the ICD9 ambulatory diagnoses recorded for them over a period of time. The methods used to assign ACGs, one of the models tested in this study, are described in greater detail in Chapter III.

DCGs. DCGs were developed by Arlene Ash, Randy Ellis, and others, mainly at Boston University (Ash, Porell, Gruenberg et al., 1989; Ellis et al., 1995), with the intention of considering the implications for future medical need of different types of observed medical expenditures. In their original form, DCGs relied on inpatient diagnoses (ICD9) in one year to predict total expenditures in the next year. In doing this, ICD9 codes are grouped not according to clinical similarity, but according to similarity of predicted costs for the subsequent year. In grouping inpatient diagnoses, care was taken to distinguish hospitalizations deemed to be "discretionary" in nature from other hospitalizations.

The DCG method has undergone a number of revisions since its inception in the late 1980s, to refine its treatment of information on inpatient stays and also to make use of ambulatory diagnoses. These versions of the DCG model are among the models we tested and are described in Chapter III.

Robinson-Luft Multiequation Model. This method was developed by James Robinson at Berkeley and Harold Luft at the University of California–San Francisco and uses demographic characteristics commonly available in the personnel files of large employers, including age, sex, family status, education, occupation

and salary level (Robinson, et al., 1991). The unique feature of this model is its use of a six-equation model to estimate relative risk. This approach is based partly on the multiequation models used to analyze data from the RAND Health Insurance Experiment and includes equations to distinguish users versus nonusers of service and the level of expenditures of those using services.

PACS. PACS were developed by a team led by Gerard Anderson at Johns Hopkins University (Anderson, Cantor, Steinberg, and Holloway, 1986). This is a combined demographic and prior use model with health status measured as a combination of age, sex, disability status, and three variables that define prior use (major diagnostic category associated with each hospitalization, chronicity of each disorder, and ambulatory resource use).

Self-Reported Health Status. A variety of self-reported health status measures have been used for risk assessment. These include measures based on the Rand 36-Item Health Survey, a questionnaire that provides a generic measure of patient functioning and well-being. The Short Form-36 (SF-36) is designed to capture multiple dimensions of health, and to measure the full range of health states, including levels of well-being. The survey is self-administered and has been widely used in outcome studies. Other scales measuring physical, mental, and general health have also been evaluated (Thomas and Lichtenstein, 1986a; Newhouse et al., 1989). The number of workdays missed in the previous four weeks and whether the person is a smoker, etc., have also been proposed (Hornbrook, Goodman, Bennett and Greenlick, 1991).

Physiologic Health Measures. Physiologic measures collected as part of the RAND Health Insurance Experiment have also been evaluated as a risk assessment measure. These include cholesterol level above 259, diastolic blood pressure above 89, glucose level above 159, and other measures (Newhouse et al., 1989; Howland et al., 1987). Such measures obviously go beyond what is normally available from claims records.

Mortality. Health plans that experience higher-than-average mortality will tend to experience higher costs, particularly so when the causes of death fall into certain high-cost groups (for example, end-stage renal disease, cancers). The suggestion has therefore been made to risk-adjust across plans based on their mortality experience (Tolley and Manton, 1984).

Health Service Use. Some researchers have proposed basing risk assessment directly on health service use, including total expenditures (Wouters, 1990; Thomas and Lichtenstein, 1986b), and number of hospital days in the past two years (Beebe et al., 1985). A related proposal has suggested using a blend of prospective risk assessment (ACGs, DCGs, or some other measure) and health service use, where payment for health service use is based on predetermined rates such as DRG hospital price or resource-based relative value schedule (RBRVS) physician prices. To provide incentives for efficiency in the provision of care, payment rates for use would be rescaled to reflect the marginal, or incremental, costs of providing the service (Newhouse, 1994).

2. Criteria for Evaluating Risk Assessment Methods

A practical and effective method for risk assessment must meet several criteria:

- It must predict expenditures well enough so that insurers do not engage in risk selection to any meaningful extent and insurers are compensated equitably for the risks that they enroll.
- It must be administratively feasible.
- It must be resistant to gaming behavior by insurers.
- It must not give insurers or providers incentives to engage in socially uneconomic behavior, such as unnecessary hospitalizations.

Of these four criteria, the first has received the most attention in the literature. Two issues in particular have been discussed: (1) is predictive accuracy needed at the level of the group, or of the individual? and (2), how well must a risk adjustment method be able to predict expenditures? We focus this review on predictive accuracy. We discuss the other evaluation criteria in detail in Chapter VI.

Group-Level Versus Individual-Level Prediction. In considering whether group- or individual-level predictive ability is required, researchers have noted that, from the perspective of the insurance plan, only the ability to predict expenditures well for groups is necessary (Lubitz, 1987; Robinson et al. 1991; Robinson, 1993; Luft, 1995). Errors in prediction at the individual level tend to cancel each other out. Others, most notably Newhouse (1993,1994), consider that such a view "ignores the behavioral incentives of the HMO, which can make more money by discouraging enrollment (or encouraging disenrollment) of any individual whose expected cost exceeds revenue" (Newhouse, 1993, p.42). Given that, the relevant measure depends on the objective considered: whether that of achieving a fair

redistribution across plans, or of minimizing incentives for risk selection. The research in this area has explicitly recognized the importance of both objectives and evaluates risk adjustment mechanisms on their ability to predict at both the individual and the group level.

Prediction for Nonrandom Groups. A requirement of any risk assessment method is its ability to produce unbiased predictions for subpopulations of enrollees. A model which consistently underpredicts the costs of one group, while overpredicting the costs of a second group, provides clear incentives and targets for risk selection. It also creates inequities for health plans enrolling disproportionate numbers of these individuals. Persons with extremely high or low relative expenditures in previous years or those with specific conditions such as cancer or heart disease are examples of such nonrandom groups.

Level of Accuracy Needed in Individual-Level Prediction. An adequate risk assessment method does not need to predict or explain all of the variance in expenditures across individuals in order to prevent risk selecting behavior (Newhouse, 1994). Many health expenditures are inherently unpredictable—acute care following an auto accident provides an extreme example. Expenditures for an individual with a high risk of a heart attack or diagnosed with cancer may also be difficult to predict (Luft, 1995). The individual at risk for a heart attack may in fact recover fully and need minimal care for years to come; or he or she may experience another, massive heart attack and die without consuming any more health care resources; alternatively, the individual may have a second heart attack and require many days in intensive care. A patient diagnosed with breast cancer might respond very well to a first round of therapy, or she might not and occasion much higher expenditures. No plan can know such things ahead of time, let alone who is going to be diagnosed with cancer for the first time, and who will experience an auto accident. This implies that no plan can fully predict the expenditures a particular individual will account for. And if the plan cannot, then a risk assessment method does not need to either. The risk assessment method only needs to be able to predict expenditures about as well as the plan could. If it does this, so that the plan can expect to be compensated fairly for individuals it predicts will cost it more, then the plan loses its incentive to engage in risk selection.

This raises the question: how well *could* a plan predict the expenditures of its enrollees? In other words, what proportion of the variance in its expenditures could it explain? If we could determine this, then this would give us an indication of the benchmark against which risk assessment methods ought to be measured.

Several researchers have attempted to do this. Newhouse (1982) and Welch (1985) estimated that 20% of the variance in total health care expenditures was predictable using information on previous claims available to a plan. Using a calculation based on the "between" versus total variation approach, Fowles et al. (1994) obtained virtually the same figure. McCall and Wai (1983) and Newhouse et al. (1989) have found a somewhat lower figure. Newhouse recently concluded that: "The literature suggests that a plan should be able to predict at least 15 to 20% of the variance in annual spending across a random sample of the population; thus, that is the target figure for an adjustment formula" (Newhouse, 1994, p. 140).

Others have noted, however, that risk selection is not costless. It requires, at a minimum, the collection and analysis of information and may also involve practices, such as keeping patients from the highest quality care, that have the potential to harm the insurer's reputation (van de Ven et al., 1994). It is possible, therefore, that a risk adjustment mechanism need not explain quite as much as 20% of the variance. Alternatively, it can be argued that people can be induced to switch plans or choose away from a plan based on information they have, even if the plan doesn't. For example, if a plan gives subtle messages of being homophobic, gay men will avoid joining, and the plan needs collect no information on HIV status.

To what extent a plan can select risk and just how substantial the costs of risk selection are remains to be determined. However, considering the information a plan can conceivably collect and employ in selecting risk and therefore, what portion of the predictable variance a risk adjustment mechanism needs to explain, is one useful approach to considering the adequacy of existing risk assessment methods.

3. Evaluations of Risk Assessment Methods

Much of the early research evaluating risk assessment focused on the AAPCC. Although it scores very well on the dimensions of administrative feasibility, gameability, and incentives, it has very little power to predict individual expenditures. Lubitz, Beebe, and Riley (1985) showed that AAPCC adjusters, location excepted, explained only 0.6% of the individual variation in annual

Medicare-covered expenditures for the elderly. Newhouse et al. (1989), using results in Anderson et al. (1986), estimated that the addition of location to the adjusters would cause the explained variance to rise to about 1%.

Starfield et al. (1991), using data from four HMOs and a large Medicaid population, found ACGs could explain over 35% of the variation in ambulatory charges for the same year (retrospective assessment), as compared to 3 to 6% using age and sex alone. ACGs assigned in a given year explained almost 20% of this variance in the following year (prospective assessment). Weiner et al. (1991) used data from a nonelderly HMO population and found ACGs to predict retrospectively almost 40% of the variation in ambulatory charges and 15% of total charges. Prospectively, the model predicted approximately 20% of the variance in ambulatory charges.

Ash et al. (1989) used several models including an earlier version of DCGs to predict total costs for individuals and selected groups. They estimated the models using Medicare beneficiary data for 1978 and 1979 and then predicted 1980 expenditures. They found the models including DCGs to explain less than 5% of the variance in total costs across individuals in 1980. They also tested the AAPCC model and found it to perform poorly at the individual level, explaining only 0.5% of the variation in costs (even less than the estimate of 1% cited above). The DCGs performed well at predicting costs for the entire group of enrollees and for one subgroup of lower cost Medicare enrollees tested (women 65–69 years of age), yielding predictive ratios (predicted total group costs/actual total group costs) close to unity. However, when considering other nonrandom subgroups, the researchers found a DCG model would significantly underestimate for groups with intensive hospital use in the previous year and overestimate for those having no claim costs in the prior year. Using similar Medicare data for 1984 and 1985 and a somewhat refined DCG model, Ellis and Ash (forthcoming, 1995) obtained comparable results.

Using data from the Netherlands, van Vliet et al. (1994) examined the ability of different risk assessment variables to predict *future* expenditures. They found a model using age and sex could predict about 3% of the individual variation in annual medical spending. Adding an earlier version of DCGs to this model increased this amount to 6.6%. The authors noted that DCGs lose much of their predictive ability from individuals with no hospitalization in the base year but high expenditures in the prediction (second) year. On the twin hypotheses that some of these individuals might in fact be chronically ill but happen not to have experienced hospitalizations in the base year, and that some types of hospitalizations would indicate conditions with sequelae several years into the future, they are exploring using hospitalization data from additional prior years to determine the DCG assignment (multiyear DCGs). The authors also found that including chronic conditions and the number of physical impairments improved the model's predictive ability. However, under each of these models, they found significant room for health plans to potentially profit from favorable risk selection, as measured by the magnitude of the predictable variance in expenditures left unexplained by the models.

Anderson et al. (1990), using Medicare data, compared the ability of the AAPCC, cost related groups (CRGs), an earlier version of DCGs and PACs to predict prospectively total costs.[3] To do this, they estimated a risk assessment formula for each method using a sample of Medicare beneficiaries and tested the methods using other groups of 5,000 beneficiaries. They found all four methods to predict group expenditures well—the predictive ratios for the methods all being close to unity. When examining four subgroups of beneficiaries (females 66–69, cancer patients, heart patients, and patients with multiple hospitalizations) they found the predictive ratios for PACs to be closest to unity for all subgroups, although PACs underpredicted the costs of cancer patients by about 20%. Finally, they found PACs to best explain the variation in costs across individuals (30 to 35%), followed by CRGs, DCGs, and the AAPCC.

Robinson et al. (1991) evaluated their six-equation demographic model in terms of its ability to predict expenditures for groups of various sizes, using data from a large employer. They found their model to predict well for groups, particularly for groups of 1,000 or more. For groups above this size, the predictive ratios were all close to unity. They concluded additional predictors such as prior use are required to reduce prediction error for smaller groups. They did not evaluate the ability of their model to predict expenditures at the individual level, but the demographic, nonclinical nature of the variables used indicates that it would not be high. Their model relies on data routinely found in employer personnel files, though not necessarily available to insurers (for example, salary level). Due to the purely demographic nature of the predictor variables, a risk adjustment mechanism based on their model would be somewhat impervious to gaming and would not provide undesirable incentives.

Using a large database from the RAND Health Insurance Experiment, Newhouse et al. (1989) examined how well demographic characteristics, health status, and prior utilization could predict future annual medical expenditures among nonelderly persons. They found that the AAPCC could explain 1.5% of the individual variation in expenditures. The addition of physiologic health measures and self-reported health status variables together brings the figure to almost 5%. These measures, however, would be costly to obtain and audit in practice. If instead prior use (outpatient and inpatient expense, separately) is added to the AAPCC, the variance explained rises to 6.4% of total variation. The use of expenditures from a prior year in a risk assessment and adjustment system, however, reduces the incentive to manage care effectively.

Fowles et al. (1994) compared demographics, self-reported health status, behavioral risk factors, chronic diseases, ACGs, and ADGs[4] in terms of predictive accuracy as well as administrative feasibility. Their analysis is unique in two important ways: it is the first to make a side-by-side comparison, on the same population, of survey-derived measures with those using administrative claims data; and it includes the most detailed and comprehensive assessment of administrative feasibility issues yet performed. At the individual level, self-reported survey data and claims-based data (ADGs) were found to perform about equally well in prospective analyses, predicting 12 and 11% of the variance in total expenditures. In retrospective analyses, ACGs and ADGs perform much better than survey data—most likely due to their stronger link to actual use. At the group level, all methods perform about equally well for completely random groups, whereas claims-based and survey-based measures both perform better than demographics alone, and each about as well as the other for nonrandom groups chosen for their high or low expected risk.

In terms of administrative feasibility, the authors report that costs of both patient surveys or claims-based systems are comparable if samples are used for surveys and claims-based systems are in place. How gameable surveys are is unclear; claims-based data are subject to upcoding but this can be monitored. The authors do not discuss the incentive effects of the different methods.

In a study that was a direct precursor to this one, the Health Insurance Association of America (HIAA, 1994) compared the predictive accuracy of age, ACGs, an earlier version of DCGs, the New York list of medical conditions, and another short list of high-cost medical events based on selected types of hospitalizations. The study also evaluated ACGs and DCGs in combination, and age and high-cost events in combination. The relative performances of these different risk adjusters were evaluated using a large database (about 750,000 lives), containing 1991 and 1992 data from eight insurance carriers. These data fell into 15 pools of varying sizes. Unlike in most other studies, performance was evaluated by the extent to which risk adjustment was able to bring net payments to the pools in line with their actual cost experience. They found that all the methods were relatively successful at bringing payments for *groups* more in line with costs, especially age combined with high-cost events; but results for a pool constructed to include primarily higher risk individuals were less promising, raising questions about the feasibility of adequate risk adjustment under a system of individual free choice among plans.

4. Conclusions

In summary, previous research indicates that:

- Simple models, such as age and sex or the AAPCC, can predict medical expenditures well at the group level, particularly for groups greater than 1,000.
- Models predict less well for nonrandom subgroups within a population such as those with previously low expenditures, cancer patients, or those with prior multiple hospitalizations.[5]
- For individual expenses:
  - The literature suggests that a health plan can expect to be able to predict 15 to 20% of the variation in expenditures across individuals in a future year—using information on their expenditures in previous years. This is a "target" figure for a risk assessment method.
  - However, the costs of engaging in risk-selecting behavior for insurers may allow a somewhat lower predictive accuracy to suffice.
  - Age, sex, and location can explain only about 10% of the target amount.
  - If health status measures are added, this increases to 20 to 30%.
  - Adding prior use further increases the predicted variance, at some cost in terms of gameability and incentives for efficiency.
- A large portion of the individual variance in expenditures a health plan itself could explain is not accounted for by the risk adjusters that have been studied. No system has yet been shown to adequately predict variations in individual expenditures.

END NOTES

1. The application of risk assessment methods in setting capitation payments, profiling providers, and performing research on outcomes measurement has typically focused on using age and sex and in some cases, the more diagnosis-based approaches such as ACGs and DCGs. The use of clinically based risk adjusters is particularly relevant for research applications, such as outcomes studies, where the risk assessment measure is often developed for the purposes of the particular study itself (Iezzoni, 1994).
2. For example, recent research has indicated that the inadequacy of the AAPCC and selection may cause Medicare to overpay HMOs (by a reported 6%) due to healthier seniors choosing HMO coverage (U.S. General Accounting Office, 1994; Hill et al., 1992).
3. The CRG model is based on prior utilization with some measures to distinguish discretionary from nondiscretionary hospital admissions (Anderson and Knickman, 1984).
4. ADGs are an intermediate step in defining an ACG for a person. We discuss this model in detail in Chapter III.
5. This finding is notable because of the need to risk adjust for insured groups with very favorable risks (for example, those resulting from "tough" underwriting) and insurers with high risks (for example, "insurers of last resort").