VI

Comparison of Risk Assessment Methods—General Considerations

There exist a number of general considerations beyond predictive accuracy when selecting a risk assessment method for use in risk adjustment. In particular, these issues relate to the practical side of this process, such as whether it can be implemented and administered, whether the data exist to support its application, and what types of incentives does it provide to perform risk selection, game the system, or to promote efficient health care.

We identified four general criteria for comparing methods:

1. *Practicality/administrative cost.* A risk adjustment system cannot be so complex and costly that it cannot be applied under real life circumstances. It should be understandable and somewhat straightforward to explain and apply. Furthermore, the data required for the approach must be readily available or feasible to obtain or develop at a reasonable cost.
2. *Ability to restrict manipulation.* A system should limit the ability of health plans to benefit financially by "gaming" the system. The model should rely on data that are objective and easily validated.
3. *Timeliness and Predictability.* In setting premiums, health plans should be able to predict with some accuracy the amount of the carrier transfer and how it will affect their premiums.
4. *Incentives for efficiency and quality care.* A risk adjustment system should provide no disincentive for providing efficient and high-quality care.

We compare each of the methods based on these principles below.

A. Practicality/Administrative Cost

The age and sex model is the most practical of the methods evaluated. Most health plans currently collect demographic information on each enrollee. The data are readily available and only require updates for those individuals leaving or joining a plan. One potential data problem with demographic models is the plan's ability to maintain the required information at the enrollee, rather than subscriber, level. However, if present, this problem creates even greater difficulties for the diagnosis-based methods.[1]

Each of the ACG and DCG models compared have similar characteristics related to these practical criteria. All of these models require a listing of the individual's prior diagnoses coded using the ICD9-CM system. While the ACG and ADG systems rely on ambulatory diagnoses, the ADCG, EDCG, ADCGDX and EDCGDX models require both inpatient and ambulatory information. The PIPDCG model requires only inpatient data.

In contrast to age and sex, all of the diagnosis-based models would be more expensive to administer, initially and on an ongoing basis. Costs would vary depending on whether plans currently collect and maintain these data. Since the PIPDCG model requires only the collection of inpatient data, costs would be expected to be relatively lower for this approach than the other diagnosis-based methods.

In terms of understandability, all of the diagnosis-based systems involve complex algorithms which group diagnoses for assignment to ACGs, ADGs or DCGs. However, the grouping algorithms all follow a systematic logic and could be made available to all

plans. With a modest investment of time, all plans should be able to gain some understanding of each of the systems. The ACG and DCG researchers' emphasis on an "open-architecture" in developing and applying their models should help in this regard.

Of the diagnosis-based models, those relying on ambulatory diagnoses are less practical from the perspective of data availability. There does not currently exist significant uniform or centralized collection of the ICD9 codes associated with ambulatory encounters. Some carriers, including Medicare and selected national carriers, are the exception and employ common standards in recording this information. However, given the potentially large number of insurance carriers participating in a risk adjustment system, and the likely diversity in the sophistication of their data systems, significant problems could exist in obtaining both consistent and high quality information on ambulatory diagnoses.[2]

Ambulatory diagnoses may not be readily available for other reasons. For example, some HMOs use salaried physicians or capitated payments and do not maintain ambulatory care encounter data. Further, ambulatory diagnoses for those individuals uninsured in a previous period or insured in another region of the country not participating in the risk adjustment system may be difficult to obtain, an issue particularly relevant for prospective risk assessment applications.

In contrast, inpatient data are more generally available and are maintained by nearly all health plans, including HMOs. These data are typically recorded in a more uniform way and are often maintained by state entities for different purposes such as rate regulation (for example, Massachusetts, New York, and Maryland).[3] While ambulatory data are recorded primarily by physicians and physicians' offices using various guidelines, inpatient diagnoses are recorded by hospitals using well-developed industry standards. As a result, PIPDCGs, which rely solely on inpatient data, have some advantage here.

The accurate application of the diagnosis-based models in a risk adjustment transfer process requires consistent data recording across *all plans* involved. In particular, if plans differ in their ability to capture all of the data required or in the "intensity" of their coding practice, as measured by the use of three- versus four- or five-digit level ICD9 codes or the typical number of diagnoses recorded per encounter, the assignment of individuals to risk groups can be affected. Consistency in coding may be most important to the ACG and ADG models where additional diagnoses might lead to a greater number of ADG assignments and potentially a higher ACG. The DCG models are somewhat less sensitive to these differences, given that they assign each enrollee to only one DCG, based on the single, highest cost DCG identified. However, the high-cost coexisting conditions used in some of the DCG models (ADCGDX and EDCGDX) are additive in nature (an individual can be assigned to multiple conditions) and also have a potential for variations in risk assignments with variations in coding practice.

Our simulation of a risk adjustment transfer process described in Chapter V underscores the need for consistent and complete data for applying the diagnosis-based methods. We found the estimated transfers between pools to differ greatly between those methods employing ambulatory diagnoses and the other methods used in this particular analysis. The results using age and sex or PIPDCGs were markedly different than those obtained using the ADG or the EDCGDX models. These differences were traced to the likely ability of the plans to capture claims data from all ambulatory encounters and possibly the propensity of the plans to provide a different mix of preventive and some primary care services.

In general, we found the diagnostic codings included in the study data to be of reasonable quality and intensity (as measured by the high proportion of valid codings and the relationship between expenditures and availability of codes for an individual). Differences in ambulatory coding were observed for some plans.[4] As might be expected, the inpatient diagnostic information was in better shape than the ambulatory codings. Given the requirements of the carriers for participation in the study (they had to at least have some of the information available), these carriers may be above average in terms of maintaining the data required to support these methods. We discuss the data issues for risk assessment and risk adjustment in greater detail in Appendix B.

In summary, there are clear practical advantages to the use of age and sex and other demographic models in performing risk assessment. All of the diagnosis-based models face potential problems in terms of the availability and consistency of the data used to group enrollees. Of these models, the PIPDCG method has some advantage, given its reliance on only inpatient data. Methods relying on ambulatory diagnoses face more significant challenges before they can be used in a risk transfer process.

Despite this assessment, the further outlook for the data used by the diagnosis-based models is good. In response to the demands of large employers and consumers and in an effort to understand and analyze their

expenditures on health services, plans are moving toward improving data collection capabilities. Further, any major health reform effort will likely require the development of a uniform national data collection system for both inpatient and outpatient events (Fowles, et al., 1994). These trends would be expected to enhance greatly the availability and quality of the data required to support the use of an ACG or DCG model in performing risk assessment.

B. Ability to Restrict Manipulation

Age and sex information can be verified through an audit process and is thus least likely to be manipulated.

For ACGs, ADGs and DCGs, it is possible that manipulation through upcoding could occur, as has occurred with many DRG payment systems. Given the greater uniformity in recording inpatient diagnoses and the smaller number of claims, this information may be more difficult to game and is certainly simpler to audit.

Audits of coding practice could constrain gaming behavior. Further, many argue that any risk adjustment system involving competing carriers will be self-policing. If a particular carrier is consistently upcoding diagnoses for the purposes of increasing payments, other carriers are likely to observe this and complain about this practice. Although significant abuse may be easy to detect, it remains to be seen whether plans or a regulator can identify more subtle manipulation of coding information, particularly for ambulatory encounters.

C. Timeliness and Predictability

In setting premiums and negotiating contracts, plans would require sufficient information about the likely risk transfers on a timely basis. In comparison to age and sex, the diagnosis-based methods would require significantly greater time in performing the data collection and analysis required for risk assessment. Smaller differences in predictability would exist between the diagnosis-based methods.

One practical issue related to this criteria is the difference between the time period for which inpatient episodes and ambulatory encounters occur and the point at which these data are available for risk assessment. This is particularly valid for a prospective system where risk transfers are determined in advance of the rating year. In our study, we used data from 1991 to predict costs for 1992 under a prospective design. However, it is highly unlikely that data for the previous year would be available for use in setting payments for the following year. If the data were available, a more realistic evaluation of a prospective design would involve using data for a year to predict expenditures two years later. Predictive accuracy would be expected to decrease relative to our findings under such an application.

For retrospective risk adjustment, where a settlement is made following the year to be adjusted, the time lag required to perform risk assessment is still important due to issues of equity surrounding those insureds no longer in the risk pool at the time of settlement. In addition, deferral of transfer payments could raise solvency concerns for some carriers due to the slim margins in most health insurance premiums.

D. Incentives for Efficiency and Quality of Care

Different risk assessment methods can provide different incentives for both the quality of health care services provided to beneficiaries and the efficiency with which they are provided. These incentives derive from the relationship between the type and number of health services provided and the assignment of an individual to a risk group. In this way, the age and sex model can be considered to be incentive-neutral—the provision of health services is unrelated to the variables used in determining the risk group assignment (age and sex).[5]

In comparison, diagnoses result from the number and mix of medical encounters an enrollee experiences in a year, both inpatient and outpatient. Since risk group assignments (and transfer payments) under diagnosis-based models are determined by the diagnoses observed for a group of individuals, health plans may have an incentive to provide or not to provide particular services to their beneficiaries. In particular, if the presence of an additional service results in an alternative risk group assignment for an individual, and the additional risk payment (weight) attached to that group exceeds the cost of providing the service, a plan has an incentive to provide that service. If the cost of the service exceeds the additional risk payment then plans have an incentive to withhold the service.

The incentives for efficiency and quality can vary depending on whether a model is used in retrospective or prospective risk assessment. We discuss first these issues for retrospective applications.

1. Retrospective Models

Under retrospective risk assessment, the risk assigned to a plan will more closely reflect their actual claims experience. As a result, plans may have a greater incentive to increase cost or utilization in order to achieve a higher payment at the end of the year. They also have a lesser incentive, in general, to be efficient. For example, under a retrospective ACG or ADG model each service producing a different type of diagnosis could result in an additional payment.[6] This is particularly true for individuals with ACG 52 (no ADGs), where no risk weight or payment is attached under a retrospective application. A single patient encounter would result in an ADG and a positive payment.[7]

One characteristic of DCG models that reduces the incentives for inefficient use of health services is the assignment of each enrollee to only one DCG, based on the single, highest cost DCG identified. In this way, additional services beyond that related to the highest cost diagnosis have no effect on risk group assignment. The high-cost coexisting conditions used in some of the DCG models (ADCGDX and EDCGDX) are the exception, where an individual can be assigned to multiple conditions.

Alternatively, another characteristic of some DCG models may increase the incentives for inefficient care. In particular, the EDCG, EDCGDX and PIPDCG models explicitly distinguish inpatient from outpatient diagnoses when assigning risk. Since inpatient diagnoses are typically assigned to a higher cost group, these models may provide an incentive to treat patients on an inpatient rather than outpatient basis. (The ADCG and ADCGDX models do not distinguish between inpatient and outpatient diagnoses.) The PIPDCG model is most susceptible to this problem in that it only recognizes inpatient information, while the EDCG and EDCGDX models give some weight to ambulatory diagnoses. As described previously, the DCG models do exclude a significant number of lower cost inpatient diagnoses from assignment to a higher DCG. In fact, 35% of all individuals in our data with an inpatient admission were assigned to the same PIPDCG as those without an admission. As a result, the PIPDCG model provides no incentives for admitting patients with these diagnoses in order to receive a higher risk score. It is unclear to what extent these types of incentives have been completely eliminated.

2. Prospective Models

Prospective models potentially provide some of the same sorts of incentives for inefficient care as those described above for retrospective models. An additional outpatient visit or inpatient stay might provide an increased payment in the following year. However, the major difference between the incentives under retrospective versus prospective risk assessment is the link between the risk group assigned to an individual and the weight attached to that group in determining risk adjustment.

Under retrospective adjustment, the risk weights are more closely linked to the services provided, since the diagnoses and expenditures used in assigning individuals into risk groups and computing weights are from the same time period. For prospective models, this year's diagnoses are linked to next year's expenditures. As a result, the weight for a risk group is typically lower under prospective applications due to some regression toward the mean, particularly for risk groups with higher expected costs in the previous year.[8] For example, using the PIPDCG model for a pool in the study, we estimated retrospectively a risk weight of $12,760 for PIPDCG 8. The prospective risk weight for this DCG using the same set of data was $3,737. Given this, it is less likely under a prospective design that the additional risk weight triggered by a service would exceed the cost of providing that service. Since the risk assigned to a plan is less closely tied to actual claims experience, there are greater incentives for efficiency under prospective risk assessment.

In terms of quality, while there may be some incentive to do more under retrospective risk assessment, there may be some incentive to do less under prospective models. Since the risk assigned to a plan will less closely reflect actual claims experience, plans may have an incentive to restrict care for those patients where treatment costs greatly exceed risk payments received. It is unclear to what extent these incentives would influence the quality of care, however, the direction of this incentive likely favors retrospective models.

Finally, the incentives for providing efficient and high quality health care are intertwined with the predictive accuracy of a risk assessment method. If plans are to be induced to compete on efficiency and quality of health services rather than selecting risk, they need to be adequately compensated for the risks they insure. To do this, it is necessary to have reliable and accurate methods for determining risk adjustment payments to health plans.

END NOTES

1. With demographic models, the lack of information at the enrollee level can somewhat be offset by distinguishing individual subscribers from those with one or more dependents. For example, the New York state risk adjustment system uses the age and gender of the subscriber and whether the family unit is an individual or an individual plus dependents as risk factors.
2. Surveys indicate that most managed care plans should be able to produce some information on ambulatory diagnoses, with the exception of staff and group model HMOs. In general, self-insured, indemnity and PPO plans may be somewhat less likely to record these data in centralized administrative files. We could not identify any research examining the quality of the data collected or its consistency across plans. See Fowles et al., 1994, for further discussion of these issues and the general considerations in comparing risk measurement methods, including some estimates of the costs required to maintain the necessary data.
3. All diagnostic information required for PIPDCGs is contained in the Uniform Hospital Discharge Dataset, which many states and health plans currently use to record inpatient episodes.
4. As described in Chapter III, for one carrier, it was clear that insufficient ambulatory ICD9 codes had been recorded. These data were excluded from the analysis. Further, we did observe some differences in the coding of ambulatory diagnoses for selected pools in the data analyzed. These differences did have an impact on the results of a simulation of risk transfers, as described in Chapter V.
5. It could be argued, however, that the use of age and sex for risk assessment may have some impact on incentives for efficiency and quality. This link relates to the potential inaccuracies introduced by the method. For example, if plans with relatively higher risk enrollees are not reimbursed equitably under a risk adjustment process, they would be constrained to provide lower cost care, potentially of lesser quality. On the other hand, if payments to plans with lower risk populations are high relative to the risk they enroll, they may have less incentive to provide efficient care. However, under such a scenario, they would have the financial means to provide higher quality care.
6. Since common diagnoses are grouped into the same ADG, additional visits for the same problem or condition will not produce an additional risk value.
7. Of course, this problem could be remedied in part by attaching a positive weight to ACG 52, perhaps varying by age and sex. This was not a problem for the ADG model we tested, because we included age and sex in addition to ADGs. As a result, those without an ADG still received some payment related to their age and sex.
8. It is possible that the additional risk weight attached to an additional service such as a short office visit can be greater for a risk group under a prospective versus a retrospective application. Again, this is a result of regression toward the mean, where those with previously low expenditures and a risk group with low expected costs in a previous year have expenditures somewhat closer to the mean in the following year.