Directions for Further Research

Our investigations have led us to identify a number of possible areas for further research. A number of these could be accomplished using the data assembled for the study. Others would require the collection of further data.

Certain health reform proposals allow for demographic rating variation, where risk assessment and risk adjustment would be performed within bands of age and sex groups. This investigation would involve subsetting enrollee data using the demographic rating groups to be employed, and assessing predictive accuracy separately for each using methods such as those employed in this study. The models may work more or less well within different demographic classes. The sum of the results across all classes could also be compared to our present findings. We initially considered this topic for the research study described here, but time and resources did not permit us to address it in any depth. This is a research area that has not been explored previously.

Another avenue for further research that we did not have the time to develop is prospective high-risk pooling. It would be possible to examine how great are the incentives for risk selection that remain after each one of a set of plans (for example, the seven pools we used in the simulation of risk transfers) cedes to a high-risk pool the enrollees that it predicts will have the highest expenditures in the following year. There is a considerable range of possible mechanisms for such prospective high-risk pooling: pools can cede 1% or 2% of their enrollees in return for a fixed premium, they can choose to cede a larger percentage in return for a higher premium, and so on. Different risk-sharing arrangements can also be used to give plans incentives for efficient management of individuals in the high-risk pool. The plans' selection of individuals to be ceded to the pool could be simulated using a model combining diagnoses and prior expenditures. It would then be possible to examine the predictive accuracy of the risk assessment methods examined here, with individuals, random groups, and nonrandom groups, once the high-risk individuals are removed. The net transfers across plans could also be simulated. It is possible that combining high-risk prospective risk pooling with a diagnosis-based risk adjustment process would reduce incentives for risk selection to a negligible level.

The data used in this study would also support a more in-depth analysis of how the current diagnosis-based methods of risk assessment might be improved. For example, we found that the methods produce biased payments for groups of individuals with previous inpatient diagnoses of heart disease or cancer. By examining more closely the prediction errors from our analyses, other conditions where a systematic bias exists might be identified. If so, they may suggest a modification to the risk groupings employed by the model, such as the creation of a new grouping, by splitting or combining existing groups, or both. These analyses could also suggest groups that might be potential winners and losers under each method.

The list of high-cost diagnoses we developed was somewhat crude. It involved an assessment of the medical attributes of each condition by a single clinician. The grouping of diagnoses for the analysis was also based on some simplifying assumptions. Given the size of the study database and the interest in this topic at the state level, it may be worthy of future investigation. One possible exploration could combine a list of carefully developed high-cost conditions to be used retrospectively for those enrollees with these conditions, with a prospective application of one of the risk assessment models tested in this study for all other individuals. To provide incentives for efficient care, prospective prices could be attached to each high-cost condition.

Any prospective risk assessment method will require diagnoses for individuals from a previous time period. The study database we employed included two years of data for many individuals—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 immediately to be used for 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. However, to what extent and how the relative performance of the models is affected is an unanswered question. Further, if some interim adjustments are to be made retrospectively, after all data become available, an interesting and important issue is how large the adjustment needs to be.

We simulated the risk transfer process under a purchasing alliance using seven pools from the study database. Such an analysis highlighted the practical concerns in implementing diagnosis-based risk assessment methods. It also provided insight into how transfers might differ using different methods. Such a simulation might be improved by using data closer to an HIPC environment—such as small group data only, perhaps with guarantee issue requirements and standard plan designs and a more geographically homogeneous population.