Comparison of Prospective and Concurrent Results

Table 3.6 compares the performance of the prospective and concurrent risk adjustment models with recalibrated risk weights.

Table 3.6: Comparison of Performance of Prospective and Concurrent Risk Adjustment Models - With Recalibration of Risk Weights

Risk Adjuster	Type of Risk Adjuster	R-Squared with claims truncated at $100,000 for:		Mean Absolute Prediction Error with claims truncated at $100,000 for:	
		Prospective Model	Concurrent Model	Prospective Model	Concurrent Model
ACG	Diag	.140	.376	2100	1599
CDPS	Diag	.186	.418	2070	1697
DCG	Diag	.198	.547	2032	1509
Medicaid Rx	Rx	.165	.328	2062	1797
RxGroups	Rx	.185	.376	2014	1707
RxRisk	Rx	.154	.292	2091	1854
ERG	Diag+Rx	.197	.427	1983	1582

As can be seen from Table 3.6, the concurrent models significantly outperform the prospective models.

Table 3.7 shows the increase in performance between the prospective and concurrent model. In particular, the table shows the increase in R-squared between the concurrent model and the prospective model with recalibrated weights.

Table 3.7: Increase in Performance between Concurrent and Prospective Model

Risk Adjuster	Type of Risk Adjuster	Increase in R-Squared with claims truncated at:		
		$50,000	$100,000	None
ACG	Diag	.258	.236	.183
CDPS	Diag	.232	.232	.207
DCG	Diag	.341	.349	.311
Medicaid Rx	Rx	.173	.164	.126
RxGroups	Rx	.198	.191	.147
RxRisk	Rx	.151	.138	.102
ERG	Diag+Rx	.245	.230	.199

The diagnosis based risk adjusters show a larger increase in performance than the pharmacy based risk adjusters when changing from a prospective application to a concurrent application. (This is also true when the increase in performance is expressed on a relative or percentage basis, rather than an additive basis.) In general, the increase in performance for the ERGs falls in between the diagnosis based risk adjusters and the pharmacy based risk adjusters. (When the increase is expressed on a relative or percentage basis, it is strictly true that the ERGs fall in between the diagnosis and pharmacy based risk adjusters.)

Comparison of Results with Prior Society of Actuaries Study

Table 3.8 compares the prospective performance of the diagnosis-based risk adjusters in this study to the performance of the diagnosis-based risk adjusters in the prior Society of Actuaries study. Table 3.9 shows a similar comparison but for a concurrent application.

Table 3.8: Comparison of Performance of Diagnosis-based Risk Adjusters in this Study to Performance of Diagnosis-based Risk Adjusters in Prior Society of Actuaries study – Prospective Application

Study	Prospective Application: Range of R-Squared among the Diagnosis-based Risk Adjusters with claims truncated* at:		
	Low	Medium	High
Current Study	17.2% to 22.4%	14.0% to 19.8%	9.9% to 15.4%
Prior SOA Study – All Pools	6.1% to 11.2%	Not Available	Not Available
Prior SOA Study – For Three Pools	6.0% to 11.1%	6.0% to 8.7%	3.9% to 6.1%

* The claim truncation levels refer to the following: (a) for the current study, low/medium/high refer to truncation levels of $50,000, $100,000, and no truncation and (b) for the prior SOA study, low/medium/high refer to truncation levels of $25,000, $50,000, and no truncation.

Table 3.9: Comparison of Performance of Diagnosis-based Risk Adjusters in this Study to Performance of Diagnosis-based Risk Adjusters in Prior Society of Actuaries study.- Concurrent Application

Study	Concurrent Application: Range of R-Squared among the Diagnosis-based Risk Adjusters with claims truncated* at:		
	Low	Medium	High
Current Study	42.9% to 56.4%	37.6% to 54.7%	28.2% to 46.6%
Prior SOA Study – All Pools	25.2% to 42.8%	Not Available	Not Available
Prior SOA Study – For Three Pools	25.4% to 50.4%	22.4% to 47.2%	13.1% to 33.4%

* The claim truncation levels refer to the following: (a) for the current study, low/medium/high refer to truncation levels of $50,000, $100,000, and no truncation and (b) for the prior SOA study, low/medium/high refer to truncation levels of $25,000, $50,000, and no truncation.

In general, these tables seem to indicate that the performance of diagnosis-based risk adjusters has improved significantly since the prior study. However, some of this improvement in performance may be due to differences in the two studies, including the use of different risk adjusters, the use of different data sets, and some differences in study methodology. Specifically, some of the differences between the current study and the prior SOA study include:

1. The prior SOA study used only the primary diagnosis code. The current study uses the first 2 diagnosis codes for ambulatory claims and up to 15 diagnosis codes for inpatient claims.
2. A number of the risk adjusters in the prior SOA study were run using only ambulatory claims or only inpatient claims. In the current study, all of the diagnosis-based risk adjusters were run using ambulatory and inpatient diagnosis data.
3. For the prospective analysis, the prior study had about half the number of members as used in this study. This might make it more difficult to get credible risk weights when recalibrating.

It is also possible that the quality of the data in the current study is better, which provides better results.

The prior study used claim data from 1991 and 1992. The current study uses claim data from 1998 and 1999. Given medical inflation between these time periods, the two sets of claim truncation levels are roughly equivalent.

May 24, 2002

Group Level Results

General Findings

- For prospective risk assessment based on results by 1998 medical condition, the diagnosis-based risk adjusters perform better than the pharmacy-based models. (As mentioned previously, the tests based on medical conditions may be biased in favor of the diagnosis-based risk adjusters.)
- For prospective risk assessment based on results by 1999 medical condition, the diagnosis-based risk adjusters perform similar to the pharmacy-based models on an overall basis.
- For prospective risk assessment based on results by claim dollar quintile, the pharmacy-based models tend to perform similar to or slightly better than the diagnosis-based risk adjusters.
- For concurrent risk assessment based on results by 1999 medical condition, the diagnosis-based risk adjusters perform better than the pharmacy-based risk adjusters. The difference is most notable for congestive heart failure (CHF).
- For concurrent risk assessment based on results by claim dollar quintile, the diagnosis-based risk adjusters tend to perform better than the pharmacy-based risk adjusters.
- All of the risk adjusters tend to overpredict for people with below average claim levels and tend to underpredict for people with above average claim levels.
- The risk adjusters originally developed and calibrated for Medicaid populations (CDPS and Medicaid Rx) showed significant improvement in their predictive performance when the risk weights were recalibrated.
- The performance of the other risk adjusters generally increased slightly after recalibration.
- As one would expect, the concurrent models significantly outperform the prospective models.

A more detailed presentation and discussion of the study results follows. The results shown in the following tables are based on no truncation of large claims. In general, truncating the large claims tends to bring the predictive ratios closer to 1.0, as one would expect. Truncation did not appear to cause any significant changes in the overall results or general relationships among risk adjusters. The only major change that occurred due to truncation involved the predictive ratios for Medicaid Rx and RxGroups for the HIV medical condition in 1999.

Medical Condition in 1998

Table 3.10 shows predictive ratios by medical condition for the prospective model with offered weights and untruncated claims. Members have been grouped together based on whether or not they had a particular medical condition in 1998, which was determined using diagnosis codes. Table 3.11 shows similar results for the prospective model with recalibrated weights.

Table 3.10: Predictive Ratios by Medical Condition in 1998 – Prospective Model with Offered Weights – Untruncated Claims

Medical Condition	Predictive Ratio for Risk Adjuster:							Actual 1999 Claim$ PMPY
	Diagnosis based:			Rx based:			Diag+Rx:	
	ACG	CDPS	DCG	Medicaid Rx	RxGroup	RxRisk	ERG	
Breast Cancer	NA	.57	.92	.59	.80	.67	1.14	8,383
CHF	NA	.47	.86	.37	.71	.59	.76	17,692
Asthma	NA	.98	.97	.91	.87	.87	.93	4,119
Depression	NA	.82	1.02	.74	.88	.83	.84	5,773
HIV	NA	.46	.94	.60	.84	.62	.81	15,902

When using offered risk weights, CDPS and Medicaid Rx do not perform as well as the other models. Similar to the prior results, when the models are recalibrated, the performance of these two models improves the most as can be seen when the results in Table 3.10 are compared with Table 3.11 below. The performance of the other models is about the same or shows a slight improvement when the risk weights are recalibrated.

Table 3.11: Predictive Ratios by Medical Condition in 1998 – Prospective Model with Recalibrated Weights – Untruncated Claims

Medical Condition	Predictive Ratio for Risk Adjuster:							Actual 1999 Claim$ PMPY
	Diagnosis based:			Rx based:			Diag+Rx:	
	ACG	CDPS	DCG	Medicaid Rx	RxGroup	RxRisk	ERG	
Breast Cancer	.81	.84	.92	.76	.78	.73	.99	8,383
CHF	.51	.79	.85	.57	.69	.60	.75	17,692
Asthma	.96	.95	.96	.89	.91	.87	.94	4,119
Depression	.85	.91	.97	.88	.89	.87	.91	5,773
HIV	.34	.95	.91	.92	.97	.68	.91	15,902

The models tend to underpredict the aggregate claims for each of these pools of members. However, for certain risk adjusters and certain medical conditions, the predicted claims may be very close to actual e.g., the ERGs for breast cancer show a predictive ratio of .99. The models tend to perform the worst on congestive heart failure (CHF). This may reflect that CHF is a fairly expensive condition and many of these models may lump together CHF with less expensive conditions. The two lowest predictive ratios occur for HIV and CHF for the ACG risk adjuster. This may reflect that the ACGs tend to use broader categories that are not clinically specific. As a result, it

may be more difficult to accurately predict the cost levels for very severe conditions such as CHF and HIV.

The diagnosis-based models tend to perform slightly better than the pharmacy-based models. For example, when comparing related risk adjusters (i.e., risk adjusters from the same developers such as CDPS & Medicaid Rx and DCG & RxGroups), the predictive ratios for the diagnosis-based risk adjuster tend to be closer to 1.0.

Note: As discussed in Section II, measures of predictive accuracy based on medical conditions that are defined using diagnosis codes may be biased in favor of the diagnosis-based risk adjusters.

For different medical conditions, the performance of the risk adjuster models may change significantly from the general performance levels shown in the tables in this section. For a given medical condition, a risk adjuster will naturally tend to perform better on this test if it has a medical condition category that matches more closely with the definition of the medical condition used in this study.

Medical Condition in 1999

Table 3.12 shows predictive ratios by medical condition in 1999 for the prospective model with offered weights and untruncated claims. Using diagnosis codes, members have been grouped together based on whether or not they had a particular medical condition in 1999. Table 3.13 shows results for the prospective model with recalibrated weights. Table 3.14 shows results for the concurrent model.

Table 3.12: Predictive Ratios by Medical Condition in 1999 – Prospective Model with Offered Weights – Untruncated Claims

Medical Condition	Predictive Ratio for Risk Adjuster:							Actual 1999 Claim$ PMPY
	Diagnosis based:			Rx based:			Diag+Rx:	
	ACG	CDPS	DCG	Medicaid Rx	RxGroup	RxRisk	ERG	
Breast Cancer	NA	.37	.59	.40	.54	.46	.73	11,508
CHF	NA	.24	.42	.22	.41	.34	.39	25,409
Asthma	NA	.71	.68	.74	.71	.71	.71	4,675
Depression	NA	.60	.69	.60	.70	.66	.65	6,629
HIV	NA	.30	.56	.48	.70	.49	.61	16,637

As can be seen in Table 3.12, CDPS and Medicaid Rx do not perform as well as the other models when using the offered risk weights. However, when the models are recalibrated, the performance of these two models improves the most. This can be seen when comparing the results in Table 3.12 with the results in Table 3.13 below. The performance of the other models is about the same or shows a slight improvement when the risk weights are recalibrated.

Table 3.13: Predictive Ratios by Medical Condition in 1999 – Prospective Model with Recalibrated Weights – Untruncated Claims

Medical Condition	Predictive Ratio for Risk Adjuster:							Actual 1999 Claim$ PMPY
	Diagnosis based:			Rx based:			Diag+Rx:	
	ACG	CDPS	DCG	Medicaid Rx	RxGroup	RxRisk	ERG	
Breast Cancer	.52	.54	.60	.52	.53	.50	.64	11,508
CHF	.27	.39	.42	.34	.40	.36	.40	25,409
Asthma	.72	.69	.71	.72	.73	.70	.74	4,675
Depression	.65	.66	.69	.70	.71	.69	.70	6,629
HIV	.26	.59	.56	.75	.79	.55	.68	16,637

The predictive ratios for groups based on medical condition in 1999 tend to be lower than the predictive ratios based on medical condition in 1998. This is to be expected since some of the members will be newly diagnosed with a medical condition in 1999 and it is not possible to accurately predict the claim levels for such people based on their prior conditions and prior claims.

Based on Table 3.13, when related risk adjusters are compared (i.e., CDPS vs. Medicaid Rx and DCG vs. RxGroup), the diagnosis-based risk adjusters perform similarly to the pharmacy-based risk adjusters. The biggest difference in performance occurs for the HIV medical condition.

Table 3.14: Predictive Ratios by Medical Condition in 1999 – Concurrent Model with Recalibrated Weights – Untruncated Claims

Medical Condition	Predictive Ratio for Risk Adjuster:							Actual 1999 Claim$ PMPY
	Diagnosis based:			Rx based:			Diag+Rx:	
	ACG	CDPS	DCG	Medicaid Rx	RxGroup	RxRisk	ERG	
Breast Cancer	.81	.82	.88	.72	.77	.67	.91	11,508
CHF	.56	.90	.92	.53	.60	.53	.67	25,409
Asthma	1.03	1.12	.95	.86	.87	.83	.90	4,675
Depression	.94	.91	.94	.85	.86	.83	.84	6,629
HIV	.49	1.11	.99	1.09	1.11	.64	.85	16,637

As one would expect, the concurrent model performs significantly better than the prospective model.

Based on Table 3.14, when related risk adjusters are compared (i.e., CDPS vs. Medicaid Rx and DCG vs. RxGroup), the diagnosis-based risk adjusters perform better than the pharmacy-based risk adjusters. The performance is markedly different for the CHF medical condition.

Claim Dollar Quintiles based on 1999 Claim Dollars

Table 3.15 shows predictive ratios for each risk adjuster by claim dollar quintile. The quintiles represent groupings of members based on each member's 1999 claim dollars. Quintile 1 represents the 20% of the population that had the lowest claim dollars. Quintile 5 represents the 20% of the population that had the highest claim dollars. Specifically, quintile 1 had actual claim dollars per member per year of $11. Quintile 5 had actual claim dollars per member per year of $8,799 (or about 4 times the overall average of $2,232).

Tables 3.16 and 3.17 show similar results except that Table 3.16 is for the prospective model with recalibrated weights and Table 3.17 is for the concurrent model.

Table 3.15: Predictive Ratios by Claim Dollar Quintile – Prospective Model with Offered Weights – Untruncated Claims

1999 Claim $ Quintile	Predictive Ratio for Risk Adjuster:							Actual 1999 Claim$ PMPY
	Diagnosis based:			Rx based:			Diag+Rx:	
	ACG	CDPS	DCG	Medicaid Rx	RxGroup	RxRisk	ERG	
1	NA	130.79	105.56	116.37	88.65	106.07	74.80	11
2	NA	8.78	6.41	8.36	5.98	6.74	5.77	194
3	NA	3.37	2.88	3.28	2.78	2.94	2.89	596
4	NA	1.59	1.65	1.64	1.68	1.68	1.76	1,560
5	NA	.40	.51	.43	.54	.49	.54	8,799
Total	NA	1.00	1.00	1.00	1.00	1.00	1.00	2,232

These results indicate that all of the risk adjusters tend to overpredict for people with below average claim levels and tend to underpredict for people with above average claim levels. The predictive ratios are quite high for quintile 1 mainly because predicted claim dollars are divided by actual claim dollars and the actual claim dollars pmpy is very close to 0. In any given year, a significant percentage of the population will have zero claims, but none of the prospective risk adjusters predict a score of zero for a given member. As a result, the prediction ratios, which have a number very close to zero in the denominator, are extremely high.

Similar to the individual level results, CDPS and Medicaid Rx do not perform as well as the other models when using the offered risk weights. When the models are recalibrated, the performance of these two models improves the most as shown in Table 3.17 below. The performance of the other models tends to be the same or show a slight improvement when the risk weights are recalibrated.

Table 3.16: Predictive Ratios by Claim Dollar Quintile – Prospective Model with Recalibrated Weights – Untruncated Claims

1999 Claim $ Quintile	Predictive Ratio for Risk Adjuster:							Actual 1999 Claim$ PMPY
	Diagnosis based:			Rx based:			Diag+Rx:	
	ACG	CDPS	DCG	Medicaid Rx	RxGroup	RxRisk	ERG	
1	92.16	98.16	80.26	88.04	82.76	97.70	68.21	11
2	6.92	6.38	6.04	6.10	5.85	6.30	5.55	194
3	3.10	2.91	2.94	2.85	2.79	2.85	2.86	596
4	1.73	1.68	1.73	1.73	1.71	1.71	1.78	1,560
5	.48	.51	.53	.52	.54	.51	.55	8,799
Total	1.00	1.00	1.00	1.00	1.00	1.00	1.00	2,232

For the prospective model, when comparing related risk adjusters, the pharmacy-based models tend to perform similar to or slightly better than the diagnosis-based risk adjusters. For the concurrent model, shown in Table 3.17 below, the diagnosis-based risk adjusters tend to perform better than the pharmacy-based risk adjusters.

Table 3.17: Predictive Ratios by Claim Dollar Quintile – Concurrent Model with Recalibrated Weights – Untruncated Claims

1999 Claim $ Quintile	Predictive Ratio for Risk Adjuster:							Actual 1999 Claim$ PMPY
	Diagnosis based:			Rx based:			Diag+Rx:	
	ACG	CDPS	DCG	Medicaid Rx	RxGroup	RxRisk	ERG	
1	13.96	29.42	6.58	34.92	28.91	50.77	6.41	11
2	3.48	3.22	3.15	4.06	3.82	4.42	3.00	194
3	2.32	2.06	2.19	2.30	2.23	2.32	2.21	596
4	1.64	1.57	1.53	1.63	1.58	1.62	1.66	1,560
5	.73	.74	.77	.69	.72	.66	.75	8,799
Total	1.00	1.00	1.00	1.00	1.00	1.00	1.00	2,232

Table 3.17 shows results based on the concurrent model. As one would expect, the results improve significantly when compared to the prospective model. On a concurrent basis, the diagnosis-based models tend to perform better than the pharmacy-based models. On a prospective basis, the diagnosis-based models tend to perform at a similar level as the pharmacy-based models.

Claim Dollar Ranges based on 1999 Claim Dollars

Appendices B, C, and D provide information on predictive ratios by claim dollar range. Due to the size of the tables, the results are shown only in the appendices. The tables in appendices B, C, and D show that all of the models significantly overpredict the claims for people with low claim levels and significantly underpredict the claims for people with high claims levels. The prediction error is larger for people with claim levels farther from the average.

May 24, 2002

Number of Individuals Grouped by each Risk Adjuster

Percentage of Individuals Grouped

This section presents some statistics regarding the number of individuals who were grouped (i.e., assigned a risk score based on one or more medical condition or drug therapy categories) by each risk adjuster.

Table 3.18 shows the percentage of the members that were grouped for 1998 by each risk adjuster. These percentages are based on the overall dataset, which includes both the calibration and validation subsets.

Table 3.18: Percentage of Members in the Study Dataset that are Grouped by each Risk Adjuster

% of Members grouped by Risk Adjuster:						
ACG	CDPS	DCG	Medicaid Rx	RxGroup	RxRisk	ERG
79.5%	41.4%	79.5%	63.2%	71.4%	48.5%	77.9%

As shown in Table 3.18, the ACGs and DCGs group the most members into a medical condition category. Given that both group nearly 80% of the population, the ACGs and DCGs take nearly everyone that has a medical claim and assigns them to a medical condition category. For ACGs, members assigned to ACG 5100 (which is "No Diagnosis or Only Unclassified Diagnosis & Non-Users") are defined as not grouped. For DCGs, members assigned to one or more of the HCC1 to HCC136 medical condition categories are defined as grouped.

CDPS groups the fewest members into a medical condition category. This is to be expected given the design of CDPS, which focuses on major illnesses and ignores relatively minor illnesses. Although CDPS only groups about ½ the number of members as ACGs and DCGs, on a prospective basis, it performs a little better than the ACGs and almost as well as the DCGs.

The pharmacy-based risk adjusters group between 49% and 71% of the members. The highest percentage for the pharmacy-based risk adjusters (71%) is less than the highest percentage for the diagnosis-based risk adjusters (80%). This reflects that more of the members have a medical claim than have a prescription drug claim.

For this dataset, the percentage of the population that had a medical claim in 1998 is 80.1% and the percentage of the population that had a prescription drug claim in 1998 is 71.4%. Note that RxGroups includes the following catch-all categories: RxG 125 for miscellaneous, recognized NDCs; RxG 126 for ungrouped NDCs; and RxG 127 for missing NDC value. Accordingly, it is not surprising that the percentage of the people assigned to one or more RxGroup categories matches the percentage of the people that have a pharmacy claim.

Cross-tabs for Percentage of People Grouped

The following tables present cross-tabs that show the percentage of the people grouped by each risk adjuster. Each table shows a cross-tab for a combination of two risk

adjusters. Note that the columns and rows may not sum exactly due to rounding differences.

Table 3.19: Cross-tab for Percentage of People Grouped by DCG & ACG

Grouped by DCGs:	Grouped by ACGs:		
	No	Yes	Total
No	20.5%	0.1%	20.5%
Yes	0.0%	79.5%	79.5%
Total	20.5%	79.5%	100.0%

Table 3.20: Cross-tab for Percentage of People Grouped by DCG & CDPS

Grouped by DCGs:	Grouped by CDPS:		
	No	Yes	Total
No	20.5%	0.0%	20.5%
Yes	38.1%	41.4%	79.5%
Total	58.6%	41.4%	100.0%

Table 3.21: Cross-tab for Percentage of People Grouped by DCG & ERG

Grouped by DCGs:	Grouped by ERG:		
	No	Yes	Total
No	17.9%	2.7%	20.5%
Yes	4.2%	75.3%	79.5%
Total	22.1%	77.9%	100.0%

Table 3.22: Cross-tab for Percentage of People Grouped by RxGroups & Medicaid Rx

Grouped by RxGroups:	Grouped by Medicaid Rx:		
	No	Yes	Total
No	28.6%	0.0%	28.6%
Yes	8.2%	63.2%	71.4%
Total	36.8%	63.2%	100.0%

Table 3.23: Cross-tab for Percentage of People Grouped by RxGroups & RxRisk

Grouped by RxGroups:	Grouped by RxRisk:		
	No	Yes	Total
No	28.6%	0.0%	28.6%
Yes	22.9%	48.5%	71.4%
Total	51.5%	48.5%	100.0%

Table 3.24: Cross-tab for Percentage of People Grouped by DCG & RxGroups

Grouped by DCGs:	Grouped by RxGroups:		
	No	Yes	Total
No	15.0%	5.6%	20.5%
Yes	13.7%	65.8%	79.5%
Total	28.6%	71.4%	100.0%

Table 3.25: Cross-tab for Percentage of People Grouped by CDPS & Medicaid Rx

Grouped by CDPS:	Grouped by Medicaid Rx:		
	No	Yes	Total
No	30.3%	28.3%	58.6%
Yes	6.5%	34.9%	41.4%
Total	36.8%	63.2%	100.0%

Section IV. Other Considerations in Selecting a Risk Adjuster

This study compares the predictive performance of a number of common health-based risk assessment methods for employees and dependents. In addition to predictive performance, there are other considerations when selecting a risk adjustment method. This section describes some of these considerations.

The initial question to ask is whether the benefit of the higher predictive power produced by health-based risk assessment methods is worth the cost of obtaining and analyzing health data on individuals. With the rise of interest in consumer choice and need for better measures of provider and plan accountability, it is likely that accounting for selection bias will be a critical issue for modern health insurance, whether for payment, budgeting, medical management, or evaluation.

Considerations in Selecting a Risk Assessment Method

In addition to predictive performance, the selection of a risk assessment method involves other considerations, including data issues, logic for assessing risk, and implementation environment.

Data Issues

From the perspective of data used to assess risk, methods can be categorized by their reliance on demographic, prior expenditure, and/or health data, including self-reported health status. This study examines methods that use claims-based health data. These health data-based risk adjustment methods can be further divided into methods that rely on diagnosis codes from claims or encounter data, or methods that rely on prescription data as a proxy for diagnoses. Models using other health data, such as survey data on self-reported chronic disease or functional status are not included in this study.

Methods that rely on demographic risk factors, such as age, gender, and program eligibility status, are easy to administer. These methods are not measures of the care process, and therefore do not produce the incentive to change treatment or coding to maximize risk scores. Unfortunately, these methods have poor predictive value at an individual-level or for risk-skewed groups.

An individual's total prior medical expenditure is a reasonably good predictor of future expenditure. These data are easier to manage than detailed encounter data. However, the incentives related to providing care in an efficient manner are poor and the information is not useful for medical management, except in identifying high-cost cases

Health status measures, such as diagnoses and prescriptions, are good predictors and provide useful medical management information. Diagnostic data must be obtained by plans from providers. Sometimes obtaining this data is difficult for some types of plans either because the plan has a capitation contract with providers that does not require

data for payment or because they are staff or group model plans which have little or no fee-for-service experience. Ambulatory diagnoses are also somewhat unreliably coded – but the diagnostic risk assessment software available has built-in safeguards to reduce the effect of incomplete data.

Changes in coding patterns are expected. For diagnosis-based methods, the main concern is ambulatory diagnoses. Historically, these codes have not been used as the basis for payment or rate setting. When producing more diagnosis codes or when more detailed diagnosis significantly increase revenue, the potential for upcoding is real. Upcoding may occur through the discovery of new cases with a primary condition, the improved refinement of coding for severity, or the increase in the coding of all related conditions affecting treatment. The effect of upcoding creates the appearance of a higher risk population when it is compared with the population used to calibrate the prediction model. The results inflate the estimate of the total cost for a population.

Some plans or purchasers may have better access to prescription data. Prescription data are timely, relatively clean, and complete for major ambulatory drugs. In addition, these data do not need to be obtained from providers, eliminating a difficult data collection step. The incentives for efficiency may be poor if prescribing is increased in order to raise a plan or provider's risk score. Prescription-based risk assessment models generally rely on drugs believed to be non-discretionary. However, with off-label prescribing, and to the extent that discretion remains in prescribing drugs for additional diseases or for less severe or marginal forms of the disease, caution should be exercised when prescription-based models are considered for provider payment applications.

Table 4.1 compares types of risk assessment methods based on data sources.

Table 4.1: Comparison of Risk Measures

Criteria	Risk Measures			
	Demographics	Prior Expenditures	Prescriptions	Health Diagnoses
Data Quality	High	Medium	High	Medium
Prediction	Low	High	High	High
Administrative Burden	Low	Medium	Medium	High
Utilization incentive	Low/None	High	High*	Low
Diagnosis coding incentive	Low/None	Low	Low	High

* High for prescription drugs, low for all other services.

Logic in Assessing Risk

When one wishes to assess risk for a disease-specific application, the different logical approaches used by the methods evaluated in this study to produce risk scores may result in one method being better suited for the application than another.

The methods evaluated in this study differ to some extent in the number of conditions they incorporate. Some use almost all known diseases to assign risk scores. Others exclude minor, acute conditions under the assumption that these conditions are not relevant to risk selection, do not represent significant per capita costs, and including them may produce a clinically needless proliferation of these codes. If the user of risk adjustment wished to categorize all patients for an evaluation of how primary care providers are managing these frequent acute minor problems, for example, then one of the methods that includes these conditions would be preferred.

Another difference is the assignment of disease measures to risk categories. The process may produce categories that are much too heterogeneous for a specific disease of interest. Some conditions are lumped with related, yet clinically quite distinct diseases with similar costs. A disease such as diabetes, on the other hand, has its own category in most of these methods and payment is affected by coding diabetes more specifically. For other conditions, more detailed coding to describe severity will not change the assignment to a risk category beyond the simple identification of the disease.

The approach to assigning individual risk scores also varies. Some methods are additive, with additional payment made for each additional identified disease category. For payment applications, some of these categories may be arranged in hierarchies of related conditions, e.g., pulmonary conditions, with payment made for only the highest cost category in the hierarchy, the assumption being that the lower cost categories in the hierarchy indicate complications related to the more significant condition. This approach avoids "double" counting. Other methods address this relatedness of conditions by assigning individuals to mutually exclusive risk categories derived by interacting all of the individual's conditions or by identifying the individual's dominant condition.

The methods evaluated in this study have been designed to be as robust to data problems as possible while preserving predictive performance. The models typically require only one occurrence of the diagnosis or prescription in the assessment period to assign risk. The number of times the same code appears is irrelevant. Discretionary or ill-defined indicators are often excluded or assigned so as to minimize gaming incentives. This means that data need not be perfectly complete and detailed to be adequate for risk adjustment.

Implementation Environment

Another important consideration is the environmental context in which health-based risk adjustment is being implemented. One such issue, for employers especially, is the concern with access to private information. A third party may need to collect and analyze the data. In addition, the payment model may require special calibration for the specific application or population. The model may also need to be updated frequently because the relationship between the risk measure and medical expenditures may change rapidly, e.g., as prescribing patterns and the kinds of new drugs on the market change. Other factors may also need to be considered. If other major purchasers are using a particular approach, it may be less confusing to the market if the same approach is used. If multiple management uses are to be made of the risk assessment, then the use that most greatly distinguishes the performance of candidate risk assessment methods may dominate the decision. Finally, the cost of licensing and maintaining the software should be taken into account. Prices vary and some are in the public domain but may require additional outside consultants for successful implementation.

May 24, 2002

Uses of Health-based Risk Adjustment

There are many uses for health-based risk adjustment by purchasers and plans. When selecting a health-based risk adjustment method, two features differentiate the applications.

1) Does the application involve payment to providers or plans?
2) Does the application's perspective focus on targeted sub-populations, or is it global?

Using the two distinguishing characteristics, specific applications can be categorized for the following four uses:

Provider or Plan Payment / Global Perspective

These uses include health plan premium rate setting and provider capitation. Under these conditions any of the diagnosis-based methods may be preferred because they are good predictors and may introduce less of a gaming incentive than the prescription-based models. Risk selection at the provider level is usually more extreme than risk selection across health plans. When capitation or volume target incentives are used to pay providers, the concern with diagnosis gaming and over-treatment become important. The use of actual utilization data, such as prescriptions, to indicate a disease and increase payment should be avoided or approached with caution. Diagnosis data is not immune from gaming but criteria exist for diagnosing many, if not most, major conditions and this helps provide a basis for validation. An additional benefit of using health-based risk adjustment for capitation is that providers are strongly incented to provide the data.

Provider or Plan Payment / Targeted Perspective

These uses include setting disease management payment levels, e.g., carve-outs, high cost case management, or disease-specific payments. While limiting the selection to diagnosis-based models to avoid perverse incentives, one would need to explore which of the methods best captures the severity and complications associated with managing a specific disease on one hand with high cost complex cases with many co-morbidities on the other. It may also be true that for the diseases of interest, one could become satisfied that the prescription indicating the presence of the condition or its severity is non-discretionary, and then prescription-based systems or a combination of systems may be considered.

No Provider or Plan Payment / Global Perspective

These uses include setting defined contribution levels for employers and employees, provider efficiency profiling, total medical cost forecasting, and budgeting. Any of the methods could be applied for these uses because secondary incentives are weak when payment is not involved. Other factors, such as data cost and other uses for the risk assessment information would dominate the selection. A relatively new use of health-based risk adjustment in rate setting is to adjust employee premiums in defined contribution products. Different approaches to managing the extent of risk segmentation vs. pooling in new products that offer not only a choice of providers but choice among

widely differing benefit plans (including spending accounts) will probably become more important to employers as the products are implemented, and as the employers become concerned with limiting the cost to employees and dependents with serious chronic conditions.

No Provider or Plan Payment / Targeted Perspective

These uses include high-cost case identification, individual underwriting, and disease management program planning and budgeting. The selection would be based on which method is most accurate and least costly to administer regardless of data source.

Section V. Considerations in Implementing a Risk Adjuster

There are a number of additional considerations beyond selecting the risk assessment method that must be considered when implementing health-based risk adjustment. A number of lessons can be learned from purchasers and plans that pioneered the implementation of health-based risk adjustment. The implementation lessons of these purchasers have been the subject of the Risk Adjustment Impact Study, a three-year study funded by the Robert Wood Johnson Foundation, through its Health Care Financing and Organization initiative.

One component of the study described and evaluated the implementation experiences of a number of employer and public purchasers who had pioneered the implementation of health-based risk adjustment. A panel of experts, consultants with actuarial expertise who represented these purchasers, considered a number of implementation issues. Among these were lag issues, data issues, and model calibration issues. Because purchasers have not yet broadly applied prescription-based models, the following discussion addresses diagnosis-based models.

Lag Issues

One concern for implementation is the lag between the date the health problem was coded or the prescription was ordered and the use of the information for payment adjustment. Lags reduce the accuracy of the payment in two ways: (1) the length of the eligibility requirement will exclude some beneficiaries who do not meet the requirement, and (2) the longer the lag, the more the predictive power is lost. It is more difficult to accurately predict future, more distant, time periods than proximal time periods.

Lags occur in three ways, including: (1) the length of the assessment window, (2) the time required for claims and enrollment data to be available to a plan or purchaser, and (3) the time to implement the risk scoring.

The Length of the Assessment Window Required by the Model

A prospective, individual model requires continuous eligibility throughout the assessment period and for each month that payment is made. A concurrent model does not require a lengthy diagnosis history. Applying an aggregate plan-level risk score to adjust future payment eliminates the requirement that an individual be continually enrolled from the assessment period to the payment period.

The Time Required for Claims Run-out

Plans will have unequal claims run-out periods. The purchaser is unlikely to be able to influence this claims lag. It is important to allow sufficient time for all the plans to reach a similar level of data completion; otherwise payment will be biased. Prescription data may lag by only a month or less before it is available to a plan. The claim data that