Robert Wood Johnson Foundation

THE SYNTHESIS PROJECT
NEW INSIGHTS FROM RESEARCH RESULTS

ISSN 2155-3718

RESEARCH SYNTHESIS REPORT NO. 25
JULY 2013

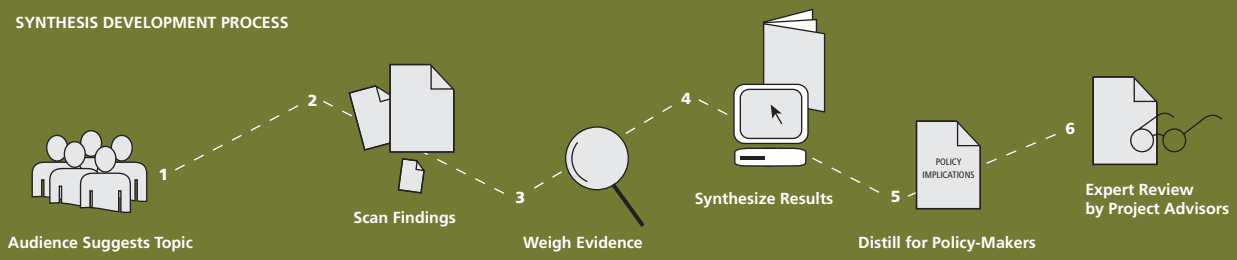
Eric Schone, Ph.D. and
Randall S. Brown, Ph.D.
Mathematica Policy Research

Risk Adjustment: What is the current state of the art, and how can it be improved?

See companion Policy Brief available at *www.policysynthesis.org*

TABLE OF CONTENTS

THE SYNTHESIS PROJECT (Synthesis) is an initiative of the Robert Wood Johnson Foundation to produce relevant, concise, and thought-provoking briefs and reports on today's important health policy issues. By synthesizing what is known, while weighing the strength of findings and exposing gaps in knowledge, Synthesis products give decision-makers reliable information and new insights to inform complex policy decisions. For more information about the Synthesis Project, visit the Synthesis Project's Web site at *www.policysynthesis.org*. For additional copies of Synthesis products, please go to the Project's Web site or send an e-mail request to *pubsrequest@rwjf.org*.



Introduction

While risk adjustment is used for a variety of purposes in the health care industry, one of its principal uses is to set payments for health plans to reflect the expected cost of providing coverage for their members. Because of differences in health status and treatment needs, the cost of health care will vary from person to person. Without risk adjustment, plans have an incentive to enroll healthier patients and avoid sick patients, especially in cases where plans cannot vary premiums based on health status, prior spending, or other known factors likely to affect their costs. Under most risk-adjustment models, plans receive a higher payment for members with multiple chronic illnesses than for members with no or limited health problems. Thus, if risk adjustment is done well, it should reduce the incentives for plans to avoid patients they expect to be costly. An accurate risk adjuster may even create incentives to attract these high-needs patients, because there may be more opportunities for the plan to provide care efficiently and generate profits. The challenge for risk adjusters is to make accurate predictions and capture all of the factors that affect expected costs.

Risk adjustment serves several critical purposes in public health insurance programs. First, by reducing profits or losses associated with enrolling members with differing risks, it helps to minimize socially wasteful competition among plans for favorable risk cases. Second, it helps plans that enroll high-risk individuals receive the resources needed to provide efficient and effective coverage. This effect is important to help prevent the "death spiral" that occurs when sicker patients flock to the plan that serves them best, resulting in the plan needing to increase premiums more than other plans, causing healthier enrollees to shift to lower-cost plans, as the plan's average risk rises. This sequence in turn leads to successive rounds of premium hikes and selective disenrollment. Third, in public health insurance programs that have both government-run and commercial options, such as the Medicare program, risk adjustment can be used to set premiums for commercial coverage that are less than or equal to the cost of government coverage for similar members. Finally, in addition to setting payments to reflect expected costs, risk adjustment can be used to adjust quality of care measures for hospitals, nursing homes, and health plans to ensure that comparisons are fair across providers.

The methodology used to risk adjust varies, depending in part on health care market regulations, the populations served and the source of payments. Risk adjustment is used in all major public programs offering health coverage in the United States–including Medicare Advantage (MA), Medicare Part D, and state Medicaid managed care programs. Risk adjustment is also a feature of national health care systems in other countries that are based on regulated competition among private payers.

Risk adjustment is a particularly important topic at this point in history because of its role in the major health reforms envisioned under the Affordable Care Act (ACA). The ACA prohibits plans in the individual and small group market–both inside and outside health insurance exchanges–from denying coverage based on pre-existing conditions or health status, eliminates medical underwriting, and limits premium variation based on age and other risk factors. As a result, risk adjustment is needed to ensure that a plan will not benefit from enrolling a disproportionate share of healthy patients. The ACA requires risk adjustment for all plans in the individual and small group market, both inside and outside health insurance exchanges. Policy-makers recognize that reliable risk adjustment is crucial for the successful functioning of state health exchanges and other insurance markets. In addition, the ACA includes value-based purchasing (i.e., paying for care based on its cost *and* quality) and public reporting of quality measures to promote quality and efficiency. To carry out these functions, risk adjustment is needed to compare providers' performance fairly and accurately on both of these dimensions.

Introduction

Regardless of whether it is for private insurance or public programs, risk adjustment that enables assessors to measure plans' or providers' actual experience relative to the expected outcome *for their specific patients* is essential to measuring performance accurately. To be successful, strategies to reduce costs without sacrificing quality depend heavily on valid risk adjustment.

While this synthesis is based largely on risk adjustment used in public programs, the lessons are applicable to private insurance as well. The range of risk-adjustment applications is broad. Therefore, to narrow the scope of this discussion to a manageable level, this synthesis addresses only the use of risk adjustment to set capitation rates or health insurance premiums. Also, since much of the experience with and research on risk adjustment is from the Medicare Advantage program, we draw most of our findings from that program but attempt to identify those that have broader applicability. The synthesis addresses the following questions:

1. How does risk adjustment work and how can effectiveness be measured?
2. What models and data are used and how do they compare?
3. What additional data sources or changes in methods have been suggested for risk adjustment?
4. How vulnerable is risk adjustment to manipulation and what can be done about it?
5. How can risk adjustment be used for newly insured enrollees?
6. How well does risk adjustment work in practice?
7. What are the greatest needs for improvement in risk adjustment and its application?

Methodology overview

This survey of risk adjustment methods and applications includes a nonsystematic, but comprehensive, review of the relevant literature as it relates specifically to the study's goals, covering the key empirical and theoretical studies relevant to study objectives. It draws on published literature, commercial research, government research, and other "grey" (unpublished, but publicly available) literature that evaluates risk adjustment techniques. In particular, we make extensive use of reviews conducted by other researchers that compare risk adjustment methods. In addition, we incorporate interviews with stakeholders from state and federal government agencies and from the health care industry about the application of risk adjustment, its shortcomings, new developments, and overall potential.

We limited our survey to rigorous studies that were published in the last 20 years and contained research relevant to the seven research questions described above. These studies drew on data from U.S. government programs and (to a lesser extent) European health care markets in which private health insurers compete to provide coverage.

Key words for our literature search included: risk adjustment, managed competition, affordable care act, Medicare Advantage, and the names of widely used risk-adjustment models.

Findings

How does risk adjustment work and how can effectiveness be measured?

Risk adjustment begins with a risk assessment, by which plan members are assigned risk scores. To perform a risk assessment, information about the enrollee that can be used to predict their costs is collected. That information includes the enrollee's demographic characteristics and medical conditions. By measuring the relationship between these characteristics and costs for a large group of enrollees, a formula is developed that can be used to calculate a risk score for each individual based on what their costs are expected to be compared with others. The impact of each factor depends on its coefficient or weight in the formula. For example, the additional amount that it costs on average to provide care for an older patient, or a patient with hypertension, is measured and added to the risk scores of older individuals with hypertension. In risk adjustment, individuals' risk scores are used to adjust the payments the plan receives to insure them.

To develop a risk adjuster, or select one risk adjuster among already available options, criteria must be established for assessing its effectiveness. The effectiveness of risk adjustment depends on the accuracy with which costs can be predicted based on enrollee characteristics. This predictive relationship is usually established by a statistical technique called multivariate regression, which produces an equation or model, measuring the statistical relationship between a patient's characteristics and that patient's cost. The technique identifies the equation that comes closest to correctly predicting costs for all of the beneficiaries in the sample. See Appendix III for a summary of risk adjustment applications.

A number of different metrics are used to determine which of several alternative risk adjusters provides the most accurate predictions of costs. The most commonly used metric is *R-square,* which estimates the proportion of variance in the outcome (in this case, cost) that is explained by the variables in a model. Increasing R-square means more accurate prediction and better risk adjustment. While R-square is a good measure of how well the model predicts for an individual, it is overly restrictive because the model needs to produce only accurate estimates for groups of individuals defined by observable characteristics, not accurate predictions for each individual's costs separately. Also, the method gives excessive weight to high-cost outliers, at the expense of making predictions for other cases less accurate. A second measure is the *Mean Absolute Prediction Error (MAPE),* which is the average absolute value of the prediction error made by the risk-adjustment model. This measure is not as sensitive to outliers as R-square. It is usually expressed as a percentage of overall average cost and a decreasing MAPE means a better model. A third measure is the *predictive ratio*, which measures the average ratio of predicted cost from the risk adjuster to actual costs, for various subgroups of the covered population. This measure addresses the main goal of risk adjusters–to ensure that the model predicts accurately for groups of people who could be identified by the plan and either sought out or avoided, depending on whether the risk-adjusted payment would tend to overpay or underpay for them relative to the expected cost. A ratio close to 1.0 means that the model predicts accurately for the group being measured. Finally, a fourth measure that is used occasionally is *predictable profit or loss*, which measures the difference between payment under the risk adjustment model and costs predicted based on the best model that can be developed, using data beyond what is included in the risk adjuster. Although predictive ratio is the superior metric for assessing model performance for specific subgroups, in reviewing the literature we rely more heavily on R-square comparisons because it is a single metric, and almost all studies report it.

Findings

What models and data are used and how do they compare?

Risk-adjustment models typically include age, sex, other demographic factors and coverage eligibility categories (e.g., enrollment in Medicaid, type of disability) as well as medical conditions identified from diagnosis codes or drugs prescribed. Table 1 identifies some of the most common models used for risk adjustment and the data on which they rely. The predictive factors to be included in the final model are identified through a combination of goodness of fit assessment using the measures described above, and clinical reasoning. Models designed in this way include Ambulatory Care Groups (ACGs), Clinical Risk Groups (CRGs), Diagnostic Cost Groups (DCGs), which includes CMS's HCC adjusters, developed by and Chronic Illness and Disability Payment System (CDPS), which are all based on encounter data, as well as similar prescription-based models such as Medicaid Rx, DxCG's RxGroups, or Ingenix Pharmacy Risk Groups (PRGs) (Table 1).

Table 1: Risk-adjustment systems

System	Diagnosis role	Treatment role	Population	Data sources	Algorithm
Ambulatory Care Groups [ACGs]	ICD-9 diagnosis codes used to classify beneficiaries	None	General	All claims for services	Mutually exclusive groupings of diagnoses based on clinical judgment and resource implications
Chronic Disability Payment System [CDPS]	ICD-9 diagnosis codes used to classify beneficiaries	None	Medicaid population—adult and child versions. Severity categories are based on resource use.	All claims for services	Groupings of diagnoses based on clinical judgment and resource implications. Beneficiaries may be assigned to multiple categories.
DxCG DCGs	ICD-9 diagnosis codes used to classify beneficiaries	None	General	Separate inpatient and all-source versions	Categories are defined on the basis of clinically coherent diagnosis groups, hierarchically combined into HCCs. Individuals may have multiple HCCs.
Impact Pro	ICD-9 diagnosis codes used to define episodes	Procedures used to define episodes	General	All claims for services, drug claims	Episodes defined on the basis of diagnosis, procedure and drug data; each member may have episodes falling into multiple categories.
DxCG RxGroups	Drug therapy category assignment may include diagnosis codes	Drug therapy categories defined from Rx claims used to classify beneficiaries	General	Prescription drug claims, may include claims for services	Drug therapy categories can be assigned to one or more categories
Medicaid RX	Diagnostic groups based on information from prescriptions	Prescriptions used to identify diagnoses	Medicaid population—adult and child versions	Prescription drug claims	Prescription drugs mapped to medical condition categories. Cost predicted based on medical condition and age/gender categories.
Ingenix PRG	Diagnostic groups based on information from prescriptions	Prescriptions used to identify diagnoses	Large managed care population, calibrated by enrollment period	Prescription drug claims	Groupings of prescription drugs mapped to diagnostic categories. Patient may be assigned to multiple categories.
Clinical Risk Groups	ICD-9 diagnosis codes	Included in category definition	General	Claims for services, may include prescription drugs	Mutually exclusive categories based on diagnostic and procedural criteria
Ingenix ERG	ICD-9 diagnosis codes and procedure codes used to define episodes	Part of episode definition	General	All claims for services, drug claims	All treatment information used in episode definition

Other models identify future costs on the basis of the type of treatment received rather than only the beneficiary's condition. These models combine information about beneficiaries' conditions and their severity as well as treatments received to identify episodes of care, such as heart attacks or coronary artery bypass surgery. These models were developed largely for profiling providers on the basis of their efficiency in managing such episodes, but the algorithms can also be used for rate setting. Examples of these types of models are Ingenix Episode Risk Groups (ERGs) and Impact Pro (Table 1). When risk adjusters include treatments, however, it is possible that treatment may be influenced by risk adjustment, because it would affect payment. To prevent this type of manipulation, diagnosis-based models are the most commonly used.

Since the 1980s, Medicare has offered managed care alternatives to its fee-for-service (FFS) benefit. Plans are paid based on expenditure patterns of similar Medicare FFS beneficiaries. When Medicare managed care plans were first offered to beneficiaries, the plans received a base rate reflecting the average Medicare expenditure per beneficiary in FFS Medicare in the enrollee's county of residence. This rate was then adjusted by a beneficiary-specific risk factor reflecting the expected future costs of the beneficiary relative to the average beneficiary. The risk adjuster prior to 2000 was based solely on age, gender, Medicaid enrollment status, and nursing home resident status. Brown et al. showed that, despite paying Medicare HMO plans 5 percent less than the risk adjuster predicted for their enrollees, these enrollees actually had expected FFS costs that were 6 percent lower than the Medicare capitated amounts the plan received (8, 9). The discrepancy arose because plan enrollees had characteristics that were associated with average FFS costs 11 percent below the expected cost predicted by the crude risk adjuster. These characteristics included fewer chronic conditions, less functional impairment, a higher self-rating of health, and a lower self-reported inclination to seek care when feeling ill, even after accounting for the factors in the risk adjuster. In addition, beneficiaries who did have costly conditions at enrollment, or developed them after enrolling, were more likely to disenroll from the plans than were healthier enrollees.

To fix this problem, CMS tested several risk adjusters. CMS selected the HCC model, a customized version of the DxCG system, due to the model's relative transparency, ease of modification and the logic of DxCG's diagnostic groupings. Beginning in 2000, payments to plans were risk adjusted using diagnostic cost groups (PIP-DCGs), a regression-based model predicting expenditures in the fee-for-service program based on conditions for which the beneficiary was treated in an inpatient setting during the prior year, along with demographic and other characteristics available from Medicare enrollment files. Starting in 2004, risk adjustment based on a prospective model using hierarchical condition categories (HCCs) and derived from the DxCG system was phased in. The current HCC model (V12) predicts Medicare FFS expenditures as a function of diagnoses for which treatment was received in any setting (not just inpatient) during the preceding year, with an R-square of about 10 percent to 11 percent, when predicting beneficiary cost variation in fee-for-service Medicare. The HCC model was developed separately for the community-dwelling aged, community-dwelling disabled, and institutionalized populations.

A revised HCC model (V21) includes refinement of diagnostic categories. For example, breast and prostate cancer were formerly grouped together but are in separate diagnostic categories for the new model (45). However, even after revisions, the optimized HCC model achieves an R-square of only about 12 percent.

Findings

In head to head comparisons, existing risk-adjustment models perform similarly in predicting future costs. None perform as well as models that include previous expenditures. Besides comparisons using Medicare experience, risk-adjustment models have been tested extensively using commercial data. The Society of Actuaries (SOA) has performed comparisons of the most prominent diagnosis-based, prescription-based and episode-based systems described above using commercial populations with several hundred thousand beneficiaries and two years of prior beneficiary history (12, 55). The two-year time frame permitted the researchers to estimate both models based on prior year claims data (prospective models) and current year data (concurrent models). They were also able to test the impact of several strategies to improve model fit: the type of risk factors included (e.g., use of prescription drugs, diagnoses, or durable medical equipment (DME) claims), inclusion of prior expenditures, the impact of claims lag, comparison of prospective and concurrent model fit, and truncation of outlier values, described in more detail below. They tested both "offered" models, that is, regression coefficients or weights developed on a benchmark population and provided with the risk adjustment software, and "recalibrated" models, which are based on coefficients or weights reestimated using SOA's test data. Table 2 summarizes the results from this testing. The table also includes a comparison of improvements in model fit resulting from truncating high-cost cases, incorporating prior expenditures and "concurrent" modeling–using as predictors the diagnoses for which the patient was treated during the same time interval as the costs to be predicted are measured.[1]

Among prospective models that use only diagnosis and demographic variables as predictors, predictive power is similar for most models tested (55). The R-square ranged from 15 percent to 17 percent and MAPE of 86 percent to 91 percent.[2] As shown in Table 2, retrospective models based on prescription drugs alone have similar explanatory power to those based on diagnoses from encounter data. Prospective models under-predict for beneficiaries with certain disease categories identified in the prior (base) year (e.g., cancer, asthma or HIV). All of the models under-predict the expenditures of beneficiaries who end up in the highest one percent of costs by about 80 percent. Though overall fit is similar, models based on prescription drugs perform more poorly than those based on encounter diagnoses when compared on the basis of predictive ratios by disease category. One possible explanation is that prescription drug-based models can identify diagnoses only by inference from prescription patterns. R-square for risk-adjustment models tested in 2002 (data from 1998 and 1999) and 2007 (data from 2003 and 2004) improved an average of 3 percentage points during that time. Possible reasons for this improvement include enhancements to ICD-9 diagnosis coding systems and coding practices, and refinements to risk-adjustment models, such as algorithms that do a better job of identifying chronic conditions and measuring their severity.

1 We did not include results for two tools that were part of SOA's test because they included prior expenditures as a predictor.

2 The CDPS tool, which is calibrated to a Medicaid population, exhibits a poorer fit, but improves substantially when recalibrated to the SOA's commercial sample. Recalibration of other models had little effect on their predictive power; thus, those results are not shown in Table 2.

Findings

Table 2: Comparison of risk-adjustment tools from the 2007 SOA study

Tool	As offered	100K truncation	250K truncation	Recalibrated + 250K truncation	Prior Costs + 250K truncation	Concurrent +250K truncation
R-square (%)						
ACGs	16.2	20.8	19.2	19.6	23.0	31.5
CDPS	12.4	17.6	14.9	17.7	24.6	36.8
CRGs	14.9	19.3	17.5		20.5	
DCGs[1]	17.4	22.3	20.6	21.3	26.5	54.5
RxGroups	16.8	23.8	20.4	20.5	27.1	36.9
PRGs	17.2	25.0	20.5	21.2	27.4	
Medicaid Rx	12.9	19.3	15.8	17.7	26.3	31.0
Impact PRO	21.3	26.3	24.4	25.6	27.2	
ERGs	16.2	23.7	19.7	20.0	26.5	43.3
Mean absolute prediction error (MAPE, %)						
ACGs	90.4	87.7	89.9	88.8	86.2	76.6
CDPS	95.8	93.4	95.3	91.9	85.6	79.6
CRGs	91.4	88.7	90.9		86.6	
DCGs	88.0	85.3	87.5	87.0	82.5	63.4
RxGroups	85.9	82.9	85.3	85.3	80.7	72.4
PRGs	86.4	83.4	85.8	85.6	80.9	
Medicaid Rx	90.2	87.3	89.6	88.4	81.9	78.0
Impact PRO	82.4	79.3	81.8	81.6	80.6	
ERGs	87.0	84.1	86.4	86.1	81.2	68.0
Predictive ratio for patients in top 1 percent of current year cost[2]						
ACGs			22.1	21.8	27.1	55.9
CDPS			14.6	18.2	24.2	51.6
CRGs			22.0		28.4	
DCGs[1]			23.4	20.5	25.2	65.3
RxGroups			19.9	18.2	24.9	51.9
PRGs			21.1	19.2	25.0	
Medicaid Rx			16.0	15.9	24.2	44.1
Impact PRO			30.0	26.9	29.7	
ERGs			17.7	18.0	24.3	54.7
Predictive ratio for patients with mental health diagnoses in prior year						
ACGs			92.3	98.0	99.2	100.6
CDPS			92.5	91.1	94.6	80.2
CRGs			89.0		101.8	
DCGs			95.9	98.5	100.3	94.7
RxGroups			88.6	89.2	96.8	79.6
PRGs			87.1	87.4	94.9	
Medicaid Rx			94.0	88.1	96.8	79.9
Impact PRO			98.0	97.8	100.0	
ERGs			91.9	92.3	97.2	85.2
Predictive ratio for patients with heart problems in prior year						
ACGs			103.1	100.3	105.6	91.3
CDPS			76.4	93.5	96.8	83.8
CRGs			99.5		115.8	
DCGs			103.2	96.0	97.0	95.8
RxGroups			89.4	95.7	97.0	75.2
PRGs			89.7	86.3	95.2	
Medicaid Rx			79.1	78.8	95.2	65.0
Impact PRO			99.8	97.0	98.6	
ERGs			92.9	94.5	97.8	89.8

1 HCCs are a specialized version of DCGs.

2 For concurrent model, cost percentile is from prior year.

What additional data sources or changes in methods have been suggested for risk adjustment?

The explanatory power of the common risk adjusters is inadequate to deter plans from "cherry-picking"—finding ways to attract patients likely to cost less than predicted and to deter those likely to cost more than the risk-adjusted capitation amount. Newhouse et al. estimated that health plans can predict 20 percent to 25 percent of beneficiaries' cost variation through participating providers or understanding the effects of their benefit design (40). This level of predictability is similar to the predictability of prior year expenditures. The authors argue that this level represents a floor that risk-adjustment applications must achieve to prevent selection from undermining managed competition. The SOA's tests suggest explanatory power in this range can be achieved only when a concurrent model is used, or when prior expenditures are included in the risk-adjustment model (55). Below, we compare the impact of these and other changes to the basic methods used in the algorithms that SOA tested. The table in Appendix IV describes the starting point, additional data source or change in modeling approach, and effect on model explanatory power when variations on the basic approach to risk adjustment were tested in the studies we reviewed.

Additional data sources provide small but meaningful increases in explanatory power. Additional data could include a longer period of data collection ("extended history collection") or new sources such as prescription drugs, self-reported health status, and laboratory or other clinical data. Additional data sources can strengthen risk adjustment if they capture information about enrollees not contained in the previous year's prescription drug or encounter records. That information may be about severity of illness within already known conditions, or it may identify conditions not reflected in the characteristics used in the base model.

Extended history can improve model fit by identifying more enrollees with chronic conditions, but those identified only from treatment in past years are likely to be less costly than those with recent treatment. When the cost of the condition is measured over a group of enrollees that includes those with less severe illness, it appears that the condition contributes less to cost. So although costs may be better predicted for the population as a whole and for those with a given condition, they may be less well predicted for the highest-cost cases. Frogner et al. and MedPAC have experimented with two and three years of data respectively and found that while overall model fit is improved by including more distant years of data, the predictive ratio for patients in the highest quantile of prior year costs is actually poorer (16, 36). This is likely because recent treatment for a condition is an indicator of severity and though more chronic conditions are identified, the less severe cases identified only from treatment in more distant years have lower costs in the current year.

Including both drug and encounter data can also identify conditions that may be missed by one source alone, but in most cases little information is gained by adding prescription data and treatment may be affected. When risk-assessment systems contain a broad array of conditions, there is little information to be gained by combining the two sources. One situation where adding the two sources together does appear to produce big gains is when the number of encounter-based and drug-based categories are small. For example, Dutch researchers added diagnostic categories based only on inpatient encounters to a small number of categories based on high-cost chronic conditions managed by prescription drugs (27). Losses from serving the highest-cost enrollees were substantially reduced by adding drugs to this combination. However, by adding prescription drugs to the array of risk adjusters, treatment choices may be influenced.

Survey data or clinical assessments can be used to identify enrollees with the most severe conditions or conditions missed by encounter records. Data on functioning are particularly important for risk adjustment for programs focused on people with disabilities and frail elders, such as states' Medicaid managed long-term-care programs and Special Needs Plans for dual eligibles. They are also used extensively in the Resource Utilization Groups (RUGs) risk adjuster that is employed in setting payments for nursing home stays.[3] In most cases, when such data are added to a broad-based risk adjuster, such as those used for Medicare Advantage plans, the improvement to overall model fit relative to ones based on diagnoses and demographics alone is small. However, predictive ratios for particular categories of patients can be improved substantially. For example, by adding information from the Children with Special Health Care Needs Screener to a predictive model of children's health care costs, Yu and Dick were able to not only improve model fit, but raise the predictive ratio for special needs children's costs from about half to almost one (56). Similarly, for the nursing home population, the RUGs risk adjuster relies heavily on clinical assessment of functioning to adequately predict costs. If survey and/or clinical data are available for all subjects, the effect on each individual's predicted cost can be included. If sampling is used to keep costs manageable, a factor based on average information from the survey can be applied to the risk-adjusted payments to reflect the cost effect of any difference between the plan's enrollees and the population on the survey measures.

Laboratory and other clinical data from medical records are potentially powerful but not yet readily available. Diagnostic-based information from claims shows which conditions a beneficiary has but relatively little about the severity of the condition, in most cases. The Comprehensive Severity Index (CSI®) model attempts to improve the measures of severity by using disease-specific, physiologic severity data collected from medical records in its inpatient, ambulatory, long-term care, hospice, and rehabilitation modules (21). While this added information can substantially improve model fit, it is expensive to collect and is generally used at this time only for research or provider profiling. In the future, as electronic health records become more widespread, and protocols or standards for data elements and interoperability are implemented, the potential value of more detailed clinical data for risk adjustment is considerable.

Changes to regression models do as much or more to increase explanatory power as do additional data sources. Changes to models that can increase their power include adding more risk factors, using a different time frame, changes to the treatment of the highest-cost cases, and including prior use or expenditures.

Additional risk factors for chronic conditions improve overall model fit slightly. In an independent look at the accuracy of the Medicare managed care payment methodology, Frogner et al. measure underpayment to MA plans for patients with chronic conditions when the HCC model is used (16). Adding specific interactions of chronic conditions and the count of chronic conditions improved the fit of the predictive model and corrected underpayments for patients with the targeted conditions, though overall model fit was improved only slightly. Furthermore, the model still substantially under-predicts the expenditures of those in the top 5 percent of prior year expenditures. In fact, the predictive ratio for the top 5 percent drops slightly.

Concurrent modeling improves model fit substantially, but does not include all the information contained in prior expenditures. Not surprisingly, concurrent models–that is, models that are estimated on diagnostic information from the year for which costs are being predicted–fit better than any prospective model by a wide margin. In SOA's testing

3 Use of home health care or skilled nursing facilities in the prior year could also serve as a measure of functional impairments; however, they would be subject to obvious gaming opportunities.

(Table 2), concurrent models offered improvement in R-square ranging from 13 to 33 percentage points and similar improvements in MAPE. Use of DxCGs as risk factors under concurrent modeling, for example, produced an R-square of 51 percent, compared with only 17 percent with prospective modeling (55).[4] Much of the improvement comes from unpredictable costs such as cancer diagnosis, strokes, heart attacks, and other event-related costs that occur during the year for which costs are being predicted. Prospective models, on the other hand, are effectively trying to predict, on the basis of *prior* encounters for conditions, the individuals who will experience such events in the coming year. The improvement in fit for concurrent modeling arises in large part from capturing the costs of treating emerging acute conditions or exacerbations of chronic conditions. These unpredictable costs do not affect the favorable selection that risk adjustment is intended to counteract, because they do not affect the profit a plan can expect to receive from covering a particular beneficiary. The predictive ratios for disease cohorts based on prior conditions shown in Table 2 are no better for concurrent models than for prospective models. Also, using concurrent models means that plans do not get credit for avoiding conditions, or avoiding the worsening of a patient's condition. Despite their substantially improved overall predictive power, all concurrent models still under-predict average cost for individuals in the top percentiles of costs from the prior year, and over-predict for those in the bottom percentiles of prior year cost.

Dudley et al. tested a hybrid concurrent-prospective model approach using SOA data (13). They found that a model based on prospective HCCs for most beneficiaries, but with concurrent HCCs for the 9 percent of beneficiaries with selected chronic conditions, resulted in an R-square similar to the fully concurrent model. Predictive ratios for selected groups improved substantially. This finding suggests that a blended model could yield nearly as accurate a model as the fully concurrent model while targeting explanatory power to reduce predictable profits and losses and only requiring updated data on a fraction of the enrollees.

Costs are easier to predict when the impact of the highest-cost cases is reduced. Highly skewed cost data—that is, the presence of high-cost outliers in a small number of cases with much higher costs than others—can distort risk-adjustment models, leading a number of researchers to propose alternatives. Truncation of costs above a threshold amount is the most commonly adopted strategy for improving model fit in the face of skewed cost data. SOA results from Table 2 show consistent improvement in model fit from truncation. R-squares for the diagnosis-based risk-adjustment models increase substantially when costs are truncated at $100,000; MAPE also improves markedly.[5] The findings are consistent with results from the earlier 2002 SOA study. When costs are truncated at $250,000, a lesser but still substantial improvement in model fit and in predictive ratios at extreme values occurs (12, 55). However, to provide a payment methodology that accounts for the exclusion of high costs from payment formulas when the risk-adjustment model is based on truncated costs, a method is needed to reimburse or reinsure plans for costs incurred for individual patients beyond the threshold amount.

Another approach to addressing the problems that skewness creates for the accuracy of a risk adjuster is to use a nonlinear regression model. Two-part modeling—using one regression to predict who has nonzero expenditures and a regression with the logarithm of the cost variable as the dependent variable for those with some expenditures—is one method. This type of method is substantially more complex than ordinary linear regression and is used primarily when working with small datasets rather than rate-setting applications.

4 55 percent when costs were truncated at $250K.

5 A risk adjustment tool, Impact PRO, which includes prior use in its algorithm, is excluded, as are "underwriting models" which include prior spending.

Including prior use or expenditures is the most powerful way to capture predictable future costs (16, 23, 55). Table 2 demonstrates the improvement in model fit that results when prior expenditures are added to the other risk factors in a risk-adjustment tool, substantially improving R-square, MAPE and ability to predict high-cost cases. Including measures of prior resource use in the risk adjuster can account for factors affecting expenditures that are not captured in the patients' medical condition or demographic characteristics. Such factors include unmeasured severity of the patient's diagnoses, taste for using health care, or providers' treatment patterns.

One reason that prior resource use is not used as a factor in risk adjustment models is concern that it would weaken incentives to contain costs and would actually penalize plans that succeeded in doing so. When prior expenditures are included in the models, a plan would be paid more in the current year for a beneficiary on whom more was spent in the prior year than for an otherwise identical enrollee on whom less was spent in the prior year.

Including prior expenditures in risk adjustment could be considered a form of partial capitation. Because prior cost is only one factor in a predictive model, the adjustment does not offer full reimbursement for these prior costs. Instead, when serving previously enrolled beneficiaries, the payment is a blend of fixed payments on the basis of the beneficiary's array of predetermined risk factors, and reimbursement of part of the previous year's costs. For new enrollees, when prior cost information is available, it provides a more accurate cost prediction than one based on risk factors alone. In either case, the payments encourage providers to take on beneficiaries with high prospective costs. Payers still have an incentive to economize in the treatment of these patients because they are not fully reimbursed for expenditures incurred. The payments for diagnostic risk factors plus the amount of reimbursement for expenditures incurred are sufficient to make these patients profitable if the plan can find ways to reduce their need for expensive services and/or negotiate lower prices for these services. In this sense, risk adjustment for prior expenditures functions as a type of reinsurance, and may be preferable to risk adjustment based on prescription drug use, for example. If drug use is generally correlated with higher expenditures for a given condition, a health plan may reap a profit by prescribing an unnecessary drug that is a signal of a high-cost condition but that is not so costly. By contrast, adjustment based on actual prior expenditures does not permit the plan directly to profit from overuse.

How vulnerable is risk adjustment to manipulation and what can be done about it?

Risk measurement may be subject to manipulation by providers or payers because their payment is increased by making the patient's risk appear to be as high as possible. That behavior is known as gaming. One type of gaming is upcoding, or recording extra diagnostic codes or codes representing the most severe and costly conditions that a patient might have. This phenomenon occurs as a natural result of the incentive created when recorded diagnoses are used to measure risk, and need not mean inaccurate description of the patient's condition. In fact, it may simply occur because payers are compensated for recording patient history as completely as possible. Because of gaming, the Medicare program uses a combination of methods to reduce the opportunity to profit from some types of manipulation, and reduces overall payments by a flat rate to account for systematic upcoding.

Risk factors and risk-adjustment systems are designed to reduce the impact of some types of gaming. For example, even if model fit could be improved by risk factors with negative coefficients, they are not generally included because plans could increase their payments by omitting these protective diagnoses from their records. Concurrent modeling is likely to be

particularly susceptible to gaming because diagnoses may be exaggerated to generate higher payments than warranted. Some risk factors used for determining payment may be affected by treatment choices, such as the choice of which drugs to prescribe when drugs are used as a risk adjuster or whether to admit the patient when an inpatient stay is a risk adjuster. In that case, risk adjustment has the effect of reducing the relative price of treatment options that make patients appear sicker, creating an incentive to use those treatments. For that reason, risk factors of this type are avoided in the HCC system, for example. In particular, because prescription drug regimens for chronic conditions are persistent, the predictive power of the risk adjustment model used for Medicare Part D could be greatly increased by incorporating information about which drugs are prescribed. Nonetheless, they are not included in the predictive model in order to avoid influencing treatment choices (26).

Upcoding causes the estimated average risk of Medicare Advantage enrollees to increase every year they are in the program. Upcoding has that impact because the data used to develop the risk adjustment model are derived from a different source than the data that are used to calculate payments. For example, the HCC model used by Medicare to risk adjust payment amounts is estimated on Medicare claims data submitted by fee-for-service (FFS) providers, but Medicare calculates capitation amounts for a given enrollee using data on enrollees provided by the MA plan. This situation creates the problem of "coding creep"—an increase from year to year in the number of chronic conditions and average severity of illness of plan enrollees coded on claims. MedPAC evaluated data from a fixed cohort of MA patients and a similar cohort of FFS patients over a three-year period, finding that the increase in measured risk among the MA enrollees was significantly greater. They concluded that the greater increase in the risk measure for MA plans occurred because the plans have an incentive to record more diagnoses than do FFS providers, under the HCC risk adjustment system. To compensate, MedPAC proposed a reduction in risk scores of approximately 3.4 percent for all MA, which was implemented by CMS (34). Unfortunately, this across-the-board approach effectively forces all plans to maximize the coding of conditions in order not to lose revenue. The upcoding of diagnoses may be exacerbated with the expansion of electronic health records, which makes it even easier to identify and record credible, but erroneous, diagnostic codes. (43).

A similar phenomenon, which is not due to gaming, may arise because beneficiaries treated under fee-for-service often receive different treatments than beneficiaries with the same condition who are treated through a managed care plan. Newhouse et al. investigated the impact of treatment under managed care on the pattern of expenditures using data from a large Medicare Advantage plan (41). They fit CMS's HCC model and compared coefficients on diagnostic risk factors with those in CMS's rate-setting model. They found that risk factor coefficients differed substantially between the two models. In addition they compared the fit of a model estimated on managed care data with that from the fee-for-service model and found a substantial improvement. These results suggest that problems of coding creep and selection in MA could be mitigated by using data from managed care plans to estimate risk-adjustment models and setting basic premiums by competitive bidding. However, data from MA plans needed to create and use such models are not readily and consistently available, and are likely to differ in quality and coding practices. Development of prices for services described in such data is another challenge. Furthermore, such an approach reduces the gain to managed care plans from finding more cost-effective ways of treating patients than are typically used in fee-for-service, which is a primary goal of MA.[6]

6 This objection does not apply, however, when the object is not efficiency relative to a fee-for-service benchmark, but rather preventing selection from undermining health insurance markets in which capitated health plans compete, such as ACA exchanges.

Findings

How can risk adjustment be used for newly insured enrollees?

Since prior claims history is not available for newly insured enrollees, risk factors must be identified through other means. Unlike the approaches described above that add to the array of risk factors and the amount of information collected from an enrollee's history, when the costs of people newly eligible for health coverage are modeled, such as individuals who become eligible for Medicare by turning age 65 or individuals who were previously uninsured, adjustment must be performed without prospective data from these enrollees.

The wide variation among new enrollees in Medicare with respect to costs highlights the need for adjusters in the MA program (16). Currently, MA risk adjustment for beneficiaries newly eligible for Medicare includes only demographic risk factors because prior year diagnostic information is generally unavailable. Under prospective risk adjustment, MA plans are thus faced with a delay in receiving appropriate risk adjustment for patients with high costs (although they may be overcompensated for newly eligible beneficiaries in general, if such enrollees tend to be healthier than newly eligible Medicare beneficiaries who do not enroll in MA plans). Currently, Medicare has no provision for retrospectively adjusting capitation amounts for new enrollees after their diagnostic information becomes available.

Concurrent risk adjusters, or a mixed concurrent-prospective hybrid, could be used for new enrollees. Chronic conditions are well suited for concurrent risk adjustment because they are likely to be of long duration and are unlikely to have arisen due to recent treatment, so they are not markers of poor care during the current year. In the SOA study, predictive ratios for beneficiaries with chronic conditions in the preceding year were similar between concurrent and prospective models.

Group risk adjustment is another solution for beneficiaries for whom prior diagnosis data are unavailable at the time of enrollment. Premiums can be risk adjusted based on post-enrollment diagnoses of previous cohorts of new enrollees. In other words, it would be assumed that the group of new enrollees in a health plan would have a similar mix of conditions to new enrollees from previous years. The average capitation rate paid for each cohort of new enrollees during their first year would then be relatively constant over time, except for inflation adjustment and any updating based on changes in the average risk level of new enrollees. This solution would not work in the first year of a program or health plan, during which adjustments from some other source would be applied.

How well does risk adjustment work in practice?

The effectiveness of risk adjustment can be measured by whether it increases plans' willingness to provide coverage to high-needs patients, whether it reduces favorable selection and overpayment, and whether it decreases the incentive of high-risk enrollees to disenroll. Much of the evidence of risk-adjustment effectiveness comes from experience with Medicare Advantage, and the use of HCC-based risk adjustment to set the premiums paid to health plans to cover Medicare beneficiaries. Evidence from other programs tends to be less rigorous, but some examples are included to indicate the generalizability of findings. The table in Appendix V summarizes what recent tests of risk-adjustment methods have to say about its effectiveness in achieving these goals.

Willingness to provide coverage

Numerous studies have shown that risk adjustment diversifies the population that can be profitably insured under Medicare Advantage (7, 38, 32. 42). Risk adjustment is expected to increase the number and riskiness of beneficiaries enrolled in managed care, because

health plans will be paid more for taking on enrollees whose medical conditions increase the cost of caring for them. Evidence from MA bears out this expectation. Brown et al., McWilliams et al., and Morrisey et al. found that the range of expected costs, the average of expected costs, and the actual average costs of new MA enrollees increased as stronger HCC risk adjustment for payment of MA plans was phased in from 2006 through 2008 (7, 32, 38).

Though Medicare Part D's risk-adjustment model is weak compared with the explanatory power of prior expenditures, the program provides coverage to the majority of eligible beneficiaries through a wide range of plans. Given the high cost of drugs, the high predictability of the cost for individuals with chronic diseases, and the community rating requirement, the Part D program almost surely attracts a disproportionate share of those with high drug costs. A so-called death spiral could occur under these conditions as plans that attract a disproportionate share of high-risk cases raise premiums, leading members with the lowest expected drug expenditures to leave the plan, prompting another round of premium increases and disenrollments, and eventually resulting in the plan's exit from the market. Despite the risks imposed by weak risk adjustment (23), the market did not experience severe adverse selection, ensuring that all but about 10 percent of those who are eligible have drug coverage and providing coverage to a broad and disparate array of beneficiaries, perhaps because low premiums make it broadly attractive. Currently, through subsidies, including the low-income subsidy and reinsurance, the government bears about three-quarters of the cost of the benefit (32). Using data from Part D claims, CMS revised its predictive models and was able to increase the explanatory power of its risk adjusters (26).

Experience from several European countries provides evidence that risk adjustment combined with regulation can help to sustain a managed competition health care system. Other markets in which risk adjustment plays an important role are those based on the principles of managed competition. In managed competition, beneficiaries choose between health plans that compete on the basis of quality and efficiency, under various government regulations. Several European nations have health care systems that promote elements of managed competition (e.g., the Netherlands, Germany and Switzerland to different degrees). The environment created by these systems differs in important respects from the environment created by the Medicare Advantage program. None of these systems includes a plan directly administered by the government, but private plans are heavily regulated. Each nation's system includes community rating, as well as contributions to plans equalized to predicted costs on the basis of risk adjustment. Because of regulation, both the Swiss and German systems leave relatively little scope for benefits variation compared with the variation permitted under MA. Swiss plans may vary only the degree of cost-sharing, while German benefits vary even less. The Dutch system has recently reformed to allow a degree of variation in benefit design more similar to that of the United States. In each of these national systems, nearly all eligible beneficiaries are afforded coverage, demonstrating the effectiveness of their combination of regulation and risk-adjusted coordination of contributions and expenditures. However, risk adjustment is being reformed and extended in order to preserve the systems. As plans assume more financial risk, the role of risk adjustment has increased.

Potential for risk selection

In spite of risk adjustment making it more difficult for plans to select favorable risks, enrollee self-selection may result in health plans enrolling primarily the most profitable enrollees. Even if risk adjustment is good enough to discourage plans from trying to identify and target the most profitable consumers, it is difficult to overcome biased enrollee self-selection in enrollment and disenrollment, when beneficiaries are free to change coverage. If beneficiary self-selection leads to MA plans having an enrollee mix that is healthier on average

than beneficiaries in FFS with the same observed chronic conditions and observed measures of illness severity, MA plans will receive capitation rates that overpay them relative to what the beneficiary would have cost Medicare if they were in FFS.

By identifying conditions that are associated with lower expenditures within a risk class and designing benefits that attract these lower-cost beneficiaries, plans can use risk adjustment to their advantage. Because health plans are compensated more highly for taking on patients with high expected expenditures on the basis of known risk factors, they have an opportunity to benefit from favorable selection among beneficiaries who would have been unprofitable if premiums were unadjusted. Plans seek to attract enrollees for whom they can earn the greatest profits–that is, those who will incur substantially lower costs than are predicted by the risk-adjustment model. For example, since the implementation of the more sophisticated HCC risk-adjustment model, the ratio of the expected cost of beneficiaries enrolling in Medicare managed care (based on the risk adjuster) to average fee-for-service costs has increased. This increase indicates that, as desired, plans are enrolling beneficiaries with health conditions that are more similar to those of beneficiaries who remain in Medicare FFS. It does not, however, necessarily mean that these enrollees' *actual* Medicare costs prior to enrolling have increased compared with the payments health plans receive for enrolling them.

To attract and retain a favorable mix of enrollees, plans can offer benefits and services and a care delivery style that are more attractive to beneficiaries with lower expected costs than to those with higher expected costs (7, 11, 15). Plans have the incentive to provide less generous coverage of services that are strongly correlated with higher total spending, and more generous coverage of services weakly or negatively correlated with total spending (15). For instance, generous ophthalmology benefits or fitness club memberships are likely to attract a lower risk population than enhanced oncology or pulmonary benefits would (11, 15). That is particularly the case for MA because the design of the program requires plans to rebate potential profits beyond a certain level to beneficiaries in the form of added benefits (or return the excess profit to Medicare). Also, some have argued that generous primary care benefits, at the expense of specialist benefits, will be most favorably received by healthy beneficiaries who expect that primary care will fill most of their needs (15).

While health plans have tools to promote favorable selection in MA enrollment even with risk adjustment, the evidence that they do so is inconclusive. Biased enrollment within MA appears to be lessened by risk adjustment, and remaining bias may be due primarily to self-selection. This evidence comes from MA, and is affected by features unique to that program, including deliberate overpayment from the government compared with FFS averaging over 10 percent during the study period (5). With overpayments at that level, plans' profits can be increased by enrolling relatively high cost beneficiaries.

The degree of favorable selection into Medicare Advantage plans following improved risk adjustment is inconclusive. The results of using the HCC risk-adjustment model to pay health plans created a test of the existence of biased enrollment when plans are faced with risk-adjusted premiums. Looking at what happened during this transition, researchers have reached different conclusions. Brown et al. found that the risk-adjusted profit increased after HCCs were first implemented in 2004, suggesting that health plans were benefiting from biased enrollment among those identified as high risk (7). However, using more data and including enrollment experience in subsequent years, Newhouse et al. concluded that risk-adjusted profits to MA plans from new enrollees did not increase significantly (42). As risk adjustment was strengthened over the years, the difference in risk-adjusted costs between those who switched into MA and those who

remained grew smaller, suggesting that risk adjustment was reducing selection overall. Because Part D drug coverage was implemented and movement between MA plans and FFS was restricted to an annual basis during this time, all of the effect cannot be attributed to risk adjustment. Rather than trying to isolate the impact of risk adjustment, MedPAC simply compared actual prior expenditures of MA enrollees with the analogous risk expenditures of beneficiaries who remained in FFS (35). They found that within almost all categories defined by HCC risk factors, prior costs of beneficiaries who switched to MA were lower than costs of those who remained in FFS, by 15 percent on average. These results, suggest that while risk adjustment may mitigate biased enrollment to some extent, substantial biased enrollment remains.

Disenrollment incentives

Risk adjustment may make health plans more willing to enroll a diverse group of members, but high-cost cases are most likely to disenroll (38, 42). Biased disenrollment has been observed in many studies of MA plans. High-cost cases may be particularly likely to disenroll under that program because health plans have the incentive to minimize the total cost of services, whereas the financial incentive for fee-for-service providers is to maximize expenditures. However, the difference may also reflect the desire of these high-cost beneficiaries to be freed from restrictions on which providers they may choose to see. Brown et al. found that even with risk adjustment, MA enrollees with poor health status were less satisfied than fee-for-service beneficiaries with the same health status, suggesting that biased disenrollment could result (7). Morrissey et al. found that the mean expenditures of disenrollees increased when risk adjustment is imposed (38). MedPAC found that, compared with beneficiaries already in FFS, risk-adjusted spending in 2008 of beneficiaries who disenrolled from MA in 2007 was substantially higher (35). Similarly, Newhouse et al. found that disenrollees had higher expected costs and resource use than enrollees who stayed with the plan (42). These results suggest that for the highest-cost beneficiaries incentives to disenroll remain after risk adjustment.

European systems have not eliminated the ability of plans to profit from biased enrollment or disenrollment. Dutch stakeholders argue that the effectiveness of the combination of regulation and risk adjustment in their system is demonstrated by the intense competition in their health care market that has resulted in relatively little premium variation and carriers experiencing losses (27). However, van de Ven and Schut argue that this situation is transitional; because of recent reforms, insurers have not yet learned to benefit from favorable selection (51). Bauhoff demonstrated that, in the German system, following reforms to risk adjustment that increased the role of diagnostic information but weakened demographic adjustment, plans appeared to favor enrollment of beneficiaries likely to incur lower costs based on demographic characteristics, specifically beneficiaries from former East Germany, where health care costs are lower (3).

The overall evidence about risk-adjustment effectiveness in *Medicaid* is inconclusive. Because Medicaid is administered at the state level, there are many different risk-adjustment packages and techniques in use in that program. Unlike the Medicare program, selection bias and risk adjustment have no bearing on whether the state saves money on Medicaid relative to fee-for-service because, in most cases, when states have managed care for Medicaid, it is mandated for all, or virtually all, Medicaid enrollees (although long-term-care services are typically not covered under the capitation arrangement and behavioral health services may be carved out). While individual plans may benefit from favorable selection and an inadequate risk adjuster, the managed care sector as a whole does not, when managed care is mandated. One area in which risk adjustment is particularly critical, given the high cost and heterogeneity of the population, is Medicaid managed long-term care. These plans cover expensive in-home and nursing home care.

States are increasingly turning to these arrangements, but there is very little published literature on the accuracy of such adjusters to guide them. The most accurate risk adjusters rely on data measuring both physical and cognitive functional status. However, the data are expensive to collect, and subject to overstatement of the degree of impairment if collected by the plans.

What are the greatest needs for improvement in risk adjustment and its application?

Risk adjustment now in use is based on more powerful predictive models compared with the models based on simple demographic factors that were used in U.S. and European programs only a few years ago. Even with more powerful risk adjustment, opportunities remain for insurers to profit by identifying and marketing to specific segments of the population, as do incentives for encouraging high-cost beneficiaries to disenroll. These opportunities increase as insurers have more information available about prior use and other beneficiary characteristics, and increased freedom to shape benefits and market selectively. Unfortunately, changes to eliminate these opportunities have perverse effects on incentives for insurers to invest in effective care for high-risk cases. Thus, addressing the shortcomings of risk adjusters will probably require modifying the way adjusters are used, or supplementing them with specialized payment methods.

Reduce underpayment for those in the top percentiles of predicted costs. Managed care plans have complained for many years that risk adjusters lead to underpayment for those predicted to be most expensive. This claim is supported by the research and creates incentives for plans to avoid high-risk patients (16, 55). One solution to this important flaw is to fit separate models for those with and without certain chronic conditions, or for other population subsets. A particularly attractive and effective version of such approaches is Dudley et al.'s proposal to use a concurrent adjuster for a small subset of high-risk patients and a prospective adjuster for other patients (13). This approach helps in part by identifying patients with certain conditions suggesting a very high likelihood of dying within six months after diagnosis (e.g., metastatic brain cancer, malignant pericardial effusion). Truncation is another technique for handling the highest-cost cases when developing risk adjusters. It is a simple and effective method of improving fit by reducing the influence of high-cost cases. However, it only works in conjunction with reinsurance. Previous studies have shown that using additional measures of health and illness collected through surveys improve model fit, particularly for the frail subset of the population which is likely to incur high costs. Adjustments based on survey results can be applied on an individual basis if surveys are administered to all, or as a population-based adjustment, if survey sampling is used. An important benefit that policy-makers can hope to receive from more accurate prediction of the cost of treating high-cost cases is to provide a stronger incentive for health plans to seek out chronically ill beneficiaries and manage their care efficiently.

Implement approaches to improve consistency between payments and predictions. In the MA application a particular problem is the over-prediction of costs that plans actually incur resulting from using fee-for-service Medicare as the basis for predicted costs. This inconsistency is due in part to fee-for-service beneficiaries' higher unmeasured need for or propensity to consume health care than the MA population. In addition, the fee-for-service claims on which the model is estimated are submitted by providers with a weaker incentive than MA plans to fully record the number and severity of their patients' chronic conditions. Allowing additional time for data accumulation and more accurate capture of chronic conditions is a partial solution. Unfortunately, the current solution of scaling all plans' payments down by a given factor to adjust for the higher average rate of increase in HCC scores among plan enrollees exacerbates the incentive for such upcoding, rather than reducing or eliminating it. Thus, a better long-term solution is needed.

For example, a plan's HCC score could be adjusted downward only if continuously enrolled members had a greater average increase in their HCC than did members in other plans or in fee-for-service. However, that kind of adjustment does not overcome the need for incentives for plans to take on beneficiaries with expensive pre-existing conditions or to retain beneficiaries who develop such conditions.

Use appropriate measures of goodness of fit to compare risk adjusters. Risk-adjustment methods should be selected on the basis of a metric that indicates the potential for health plans to profit from selection when risk adjustment is used. That metric might be the potential profit obtainable when a particular risk-adjustment method is used to generate capitation payments by a health plan using the most powerful possible predictive model to select beneficiaries. The most powerful model, which the plan could presumably estimate, could include multiple gameable measures such as prior expenditures or drugs prescribed, or politically controversial adjusters such as race or occupation, which are strong predictors but would not be included in a risk adjuster used to set payments. Using the mean predictive error from cohorts defined by this baseline model, policy-makers would be able to measure the scope for risk selection and possible benefits that could arise from it.

Use partial capitation to overcome the weakness of existing models, but balance incentives to contain costs. As shown above, including prior expenditures in a risk-adjustment model, which substantially improves model fit, is similar to partial capitation. This improvement in model fit would reduce the opportunity for favorable selection, and remove a source of profit or loss from enrollment of new beneficiaries. Research on methods for identifying the proper weight to give prior expenditures is needed, though, to create the correct incentives for health plans. One option for converting a model with prior expenditures to a method of partial capitation is incorporating prior expenditures as a regressor in the model. Another research topic would be to investigate the incentive effect of including prior expenditures, which could help guide the decision about how much weight should be given to prior expenditures in setting the payment amount if this approach were used. Our interviews and review of the literature suggest that partial capitation of this type is not often considered. In addition to the practical problems, policy-makers seem to feel that partial capitation weakens the incentives for cost containment excessively. For payers, protection from risk and excessive plan profits appears to be less than that provided by full capitation combined with full reinsurance for high-cost cases.

Use reinsurance to overcome underpayment for high-cost cases. Reinsurance and related risk mitigation strategies are easier to implement than other forms of partial capitation. These techniques may help to overcome selective disenrollment and under-prediction of high-cost cases. They are also consistent with risk-adjusted payments if high costs are truncated when estimating risk-adjustment models. A limit placed on a plan's financial liability for these outlier costs is needed if costs are truncated in developing the risk adjuster, to ensure that the risk faced by the health plan reflects the payment under the risk adjuster. Although the evidence suggests that reinsurance should have a role in application of risk adjustment, most implementations do not include it, or include it on a temporary basis.

ACA has reinsurance and risk corridors planned for only the first three years of exchange operations. Meeting the desired schedule for phase out may depend on whether other mechanisms (including risk adjustment) can adjust payments in a manner that does not lead to underpayment for patients with the highest predicted costs (1). The evidence we have reviewed suggests that without some form of enhancement, such as reinsurance or partial capitation, risk adjustment may fall short of this goal.

Improve incentives to preserve health. Risk adjustment is intended to counteract the incentive for plans to make their plan more attractive to patients who are low-risk and less attractive to patients with greater health care needs, among both potential enrollees and current members. Thus plans are paid more for caring for sicker patients. However, this type of payment also reduces the plan's incentive to preserve patients' health because plans are paid more as enrollees' health deteriorates. One technique for counteracting that incentive is to reward plans for good outcomes, or for care conforming to high standards. These payments could be based on quality profiling methods, which may or may not include risk-adjusted outcomes. Currently, this aspect of value-based purchasing is part of both MA and FFS Medicare. For MA, however, awards for quality are currently offered to 90 percent of plans, and seem to serve in part as a partial repeal of the reductions in MA payments that are part of the ACA. A more effective way to ensure that Medicare reaps cost savings from managed care would be to have payments geared to reward patients' health trajectories that are better than predicted and to pay less for trajectories poorer than predicted. Thus, while payment should be greater for patients with a higher HCC score, payments should be scaled back for plans whose average HCC score increases from the previous year for continuing enrollees by more than the local fee-for-service population. Plans whose enrollees have a reduction or smaller increase in their average HCC scores would receive an increase in the base rate. This approach has the advantage of both discouraging upcoding, as noted above, and rewarding higher-quality care.

Consider the interaction of regulations on variations in capitation and benefits offerings in addition to risk-adjustment techniques. Regulation that limits the scope of variation in coverage also limits the return to selection, and therefore prevents selection from unraveling markets even if risk adjusters are not especially powerful. However, the hope that managed competition will promote efficiency by enabling health plans to specialize on the basis of health care needs or preferences requires that variation in benefits be available. Furthermore, while benefits such as health club memberships are more attractive to healthy members, and therefore can exacerbate favorable selection, they may also be beneficial to the health of enrollees if they promote healthier behavior. Another form of regulation that may affect the need for risk adjustment is the minimum medical loss ratio requirement. The ACA requires that plans pay out in benefits 80–85 percent of what they receive in premiums, depending on the market. That requirement is also being added to Medicare Advantage regulations. Such a requirement may lower the return to biased enrollment. However, it remains to be seen how the requirement will be enforced. Loss ratio regulations require sufficient enrollment to ensure stable cost estimates. Aggregating cost calculations across geographic areas for a given plan sponsor weakens the requirement for local plans and does not address the problem of smaller plans that are not part of larger entities.

When a public option is available that serves as the default offer and provides benefits that are particularly attractive to those with the greatest need, such as in the Medicare program, the environment for risk adjusters is more demanding. In this case, increasing the power of risk adjusters by including prior resource use or expenditures as a determinant, or reducing the importance of highly accurate risk adjustment by partial capitation may be the best strategies for overcoming the deficiencies of risk adjusters that are generally available. Methods to harness the explanatory power of prior resource use while overcoming its many operational limitations are discussed in policy implications below.

Conclusion

Risk adjustment plays a critical role in many aspects of health care today. Yet despite the advances and increases in the ability of models to predict costs, risk adjusters continue to have some serious weaknesses. Specifically, they are still highly subject to gaming behavior–overstatement of the severity of a patient's condition, and designing benefits that appeal more to healthier patients within any risk class–to get around the intent of risk adjustment. Attempts to improve the adjusters have had limited success. While truncating high-cost cases or using concurrent diagnoses instead of prior year diagnoses to predict costs improve predictive accuracy, and therefore lessen plans' risk, they would not reduce plans' ability to engage in gaming behaviors. Plans may welcome truncation of high-cost cases and use of concurrent diagnoses, but may be less enthusiastic about another possible approach that greatly improves predictive accuracy–inclusion of prior costs in the risk adjuster–because the plan would receive less revenue in the next year as a result of reducing costs in the current year. They would still profit somewhat from reducing the use of expensive services such as hospitalizations, because they would not incur the cost of those services in the current year and the reduction in next year's premium due to lower costs in the current year would be less than dollar-for-dollar.

Implications for policy-makers

Our synthesis of the literature and opinions of some key stakeholders (see Appendix II) suggests several changes to policy that could improve risk adjustment while minimizing unintended and unwanted consequences. Some changes are primarily relevant to MA, but most are applicable to a broader range of risk-adjustment contexts. These changes include changes to risk adjusters but also changes that overcome the inherent weaknesses of available adjusters. These changes are intended to help achieve the laudable goals of risk adjustment, and align a plan's incentive to maximize profits with incentives to provide high-quality, cost-effective care.

Improved data for risk adjustment could enhance the predictive power of existing risk adjusters and reduce the incremental advantage of including prior resource use. These improvements include using a blend of prospective diagnoses and a small set of focused concurrent diagnoses, such as those recommended by Dudley et al. (13). This approach should minimize the drawbacks of concurrent adjusters, while capturing a significant portion of costs associated with "predictable" mortality. Also, CMS could consider using two years of prior diagnostic data in the risk-adjuster model for MA, and developing separate models for subgroups of patients defined by reasonably common chronic diseases for which there is a sizable within-group variance in unmeasured severity and costs (e.g., congestive heart failure, COPD). For some of these subgroups, including survey data on frailty and self-rating of health, particularly for special needs plans, may add enough value to justify its cost. Adding clinical data, such as lab results, to capture severity of illness would improve precision. Clinical data are routinely collected and could be mandated. The proliferation of electronic health records may help make this type of information more readily available. The challenge is to increase the efficiency and consistency of data capture, while overcoming incentives for gaming.

Include prior resource use while minimizing the disincentives it creates for improving chronic care. Prior expenditures can still improve the fit of risk-adjustment models beyond what would be achieved by including the additions above. This improvement can be obtained by incorporating prior resource use as a predictor, but in the form of a variable such as the number of "non-preventable" hospitalizations experienced in the prior year, rather than total costs. Such a variable should capture most of the effects of prior year costs on current costs, but would not penalize plans by reducing their future revenue if they reduce hospitalizations that researchers and clinicians view as potentially preventable with proper ambulatory care (see Moy, Barrett and Ho (39)).

Monitor the potential for plans to profit by finding gaps between predicted future costs and the risk-adjusted payment. Opportunities for payers to benefit from selection can be identified by estimating the strongest predictive model that can be devised, using any variables that a plan is likely to have at its disposal, and comparing the discrepancy between the mean of its predictions for a plan's enrollees with the mean generated by the risk-adjustment model used for premium adjustments (which would exclude some of the variables in the "best" model). Information from this approach can be used in research to narrow the gap by improving risk adjustment, or in developing algorithms to penalize plans for cherry-picking.

Reduce the incentive for diagnosis creep and upcoding. Penalties can be used to overcome the tendency of plans to increase the measured severity of illness of their beneficiaries. One method is to calculate the ratio of the average predicted cost (e.g., for Medicare, the HCC score) for continuing enrollees in the current year to the average for these same enrollees in the previous year and reduce the base capitation rate for plans that show higher increases in HCC scores than occur among other plans in the same geographic area (or, for Medicare, the change

among local beneficiaries in FFS). By increasing the base capitation rate for plans showing smaller increases in average HCC scores for the cohort of continuing enrollees than is observed among local fee-for-service beneficiaries, preventive care would be encouraged and upcoding penalized.

Incentivize plans to retain high-risk patients. Research has shown consistent evidence that selection is promoted by disenrollment of high-cost patients. The effect of selective disenrollment can be mitigated by calculating the ratio of average predicted cost to average risk-adjusted payment for a plan's disenrollees and comparing that with the analogous ratio for its retained enrollees and imposing a penalty on plans for which the ratio of these two ratios is above the market average.

Reduce the financial risk from extreme outliers that may be beyond the control of plans. Since predictive models are not suitable for identifying the most extreme cases, reinsurance will be needed. Thus, costs in developing risk adjusters could be truncated at a specific high value, and reinsurance provided to plans for costs above the limit. Because plans would have a reduced incentive to find cost-effective ways to treat patients once their costs exceed the ceiling, the payer may need to establish a clinical panel of relevant specialists to approve treatment plans for these extreme cases.

Appendix I References

1. American Academy of Actuaries Risk Sharing Work Group. *Comments on Proposed Rule on Standards Related to Reinsurance, Risk Corridors and Risk Adjustment.* Submitted to the Centers for Medicare & Medicaid Services, October 28, 2011. Accessed online at *www.actuary.org/pdf/health/RSWG_comment_letter_on_3R_proposed_rule_111028.pdf*

2. American College of Cardiology. "Principles to Guide Physician Pay-for-Performance Programs. A Report of the American College of Cardiology Work Group on Pay for Performance (A Joint Working Group of the ACC Quality Strategic Direction Committee and the ACC Advocacy Committee)." *Journal of the American College of Cardiology*, vol. 48, no. 12, 2006.

3. Bauhoff S. "Do Health Plans Risk-Select? An Audit Study on Germany's Social Health Insurance." *Journal of Public Economics*, vol. 96, no. 9–10, October 2012.

4. Beck K, Trottman M, Zweifel P. "Risk Adjustment in Health Insurance and Its Long-Term Effectiveness." *Journal of Health Economics*, vol. 29, no. 4, July 2010.

5. Biles B, Adrion E, Guterman S. *The Continuing Cost of Privatization: Extra Payments to Medicare Advantage Plans in 2008*. The Commonwealth Fund, September 2008.

6. Billings J, Zeitel L, Lukomnik J, et al. "Impact of Socioeconomic Status on Hospital Use in New York City." *Health Affairs*, vol. 12, no. 1, 1993.

7. Brown J, Duggan M, Kuziemko I, et al. "Will Risk Adjustment Decrease Health Care Costs? New Evidence from the Medicare Advantage Program." Working Paper, August 21, 2011. Accessed online at *www.ftc.gov/be/seminardocs/111215duggan.pdf*

8. Brown RS, Clement DG, Hill JW, et al. "Do Health Maintenance Organizations Work for Medicare?" *Health Care Financing Review,* vol. 15, no. 1, Fall 1993.

9. Brown RS, Hill JW. "The Effects of Medicare Risk HMOs on Medicare Costs and Service Utilization." In *HMOs and the Elderly,* Luft HS (ed). Ann Arbor, MI: Health Administration Press, 1994.

10. Buntin MB, Garber AM, McClellan M, et al. "The Costs of Decedents in the Medicare Program: Implications for Payments to Medicare+Choice Plans." *Health Services Research*, vol. 39, no. 1, February 2004.

11. Cooper A, Trivedi AN. "Fitness Memberships and Favorable Selection in Medicare Advantage Plans." *New England Journal of Medicine*, vol. 366, no. 2, 2012.

12. Cummings RB, Knutson D, Cameron BA, et al. *A Comparative Analysis of Claims-based Methods of Health Risk Assessment for Commercial Populations*. A research study sponsored by the Society of Actuaries, May 24, 2002.

13. Dudley RA, et al. "The Best of Both Worlds? Potential of Hybrid Prospective/Concurrent Risk Adjustment." *Medical Care*, vol. 41, no. 1, 2003.

14. Ellis RP. "Risk Adjustment in Health Care Markets: Concepts and Applications." In *Financing Health Care: New Ideas for a Changing Society*, Lu M and Jonsson E (eds). Weinheim, Germany: Wiley-VCH Verlag, 2008.

15. Ellis RP, McGuire TG. "Predictability and Predictiveness in Health Care Spending." *Journal of Health Economics*, vol. 26, no. 1, January 2007.

16. Frogner BK, Anderson GF, Cohen RA, et al. "Incorporating New Research into Medicare Advantage." *Medical Care*, vol. 49, no. 3, 2011.

17. Glazer J, McGuire TG. "Optimal Risk Adjustment with Adverse Selection: An Application to Managed Care." *American Economic Review*, vol. 90, no. 4, 2000.

18. Goldfield NI, Hughes JS, et al. *Adding Clinical Data to Statewide Administrative Data: Pilot Project*. Analysis and Results Final Report to the Agency for Healthcare Research and Quality, December 29, 2009.

Appendix I References

19. Hackman MB, Kolstad JT, Kowalski AE. "Health Reform, Health Insurance, and Selection: Estimating Selection into Health Insurance Using the Massachusetts Health Reform." NBER Working Paper No. 17748. Cambridge, MA: National Bureau of Economic Research, January 2012.

20. Holland P, Woolman J. *Mitigating Risk in a State Health Insurance Exchange*. A report in the Health Reform Toolkit Series: Resources from the Massachusetts Experience. Boston: Blue Cross Blue Shield of Massachusetts Foundation, December 2011.

21. Horn SD. "Validity, Reliability, and Implications of an Index of Inpatient Severity of Illness." *Medical Care*, vol. 19, no. 3, March 1981.

22. Hornbrook MC, Goodman MJ. "Chronic Disease, Functional Health Status, and Demographics: A Multi-Dimensional Approach to Risk Adjustment." *Health Services Research*, vol. 31, no. 3, August 1996.

23. Hsu J, Huang J, Price M, et al. "Distributing $800 Billion: An Early Assessment of Medicare Part D Risk Adjustment." *Health Affairs*, vol. 28, no. 1, 2009.

24. Iezzoni LI (ed). *Risk Adjustment for Measuring Health Care Outcomes, Third Edition*. Chicago: Health Administration Press, 2003.

25. Jost TS. *Health Insurance Exchanges and the Affordable Care Act: Eight Difficult Issues*. The Commonwealth Fund, September 2010.

26. Kautter J, Ingber M, Pope GC, et al. "Improvements in Medicare Part D Risk Adjustment: Beneficiary Access and Payment Accuracy." *Medical Care*, vol. 50, no. 12, December 2012.

27. Leu RE, Rutten FFH, et al. *The Swiss and Dutch Health Insurance Systems: Universal Coverage and Regulated Competitive Insurance Markets*. The Commonwealth Fund, January 2009.

28. The Lewin Group. *Medicaid Managed Care Cost Savings—A Synthesis of 24 Studies*. Report prepared for America's Health Insurance Plans. Updated March 2009.

29. MaCurdy T, Kerwin J, Theobald N. "Need for Risk Adjustment in Adapting Episode Grouping Software to Medicare Data." *Health Care Financing Review*, vol. 30, no. 4, 2009.

30. Manning WG, Mullahy J. "Estimating Log Models: To Transform or Not To Transform?" *Journal of Health Economics*, vol. 20, no. 4, July 2001.

31. McFadden DL, Noton CE, Olivella P. "Remedies for Sick Insurance." NBER Working Paper No. 17938. Cambridge, MA: National Bureau of Economic Research, March 2012.

32. McWilliams JM, Hsu J, Newhouse JP. "New Risk Adjustment System Was Associated with Reduced Favorable Selection in Medicare Advantage." *Health Affairs*, vol. 31, no. 12, December 2012.

33. Medicare Payment Advisory Commission (MedPAC). *Report to the Congress: Increasing the Value of Medicare*. June 2006.

34. Medicare Payment Advisory Commission (MedPAC). *Report to the Congress: Medicare Payment Policy*. March 2010.

35. Medicare Payment Advisory Commission (MedPAC). *Report to the Congress: Medicare Payment Policy*. March 2012.

36. Medicare Payment Advisory Commission (MedPAC). *Report to the Congress: Issues for Risk Adjustment in Medicare Advantage*. June 2012.

37. Min S-J, Fish R, Hittle D. *Trends in Risk-Adjusted Skilled Nursing Facility Rates of Community Discharge and Potentially Avoidable Rehospitalization, 2000 through 2008*. Report for the Medicare Payment Advisory Commission by staff from the Division of Health Care Policy and Research, University of Colorado. MedPAC report no. 11-3, June 2011.

Appendix I References

38. Morrisey MA, Kilgore ML, Becker DJ, et al. "Favorable Selection, Risk Adjustment, and the Medicare Advantage Program." *Health Services Research*, vol. 48, no. 3, June 2013.

39. Moy E, Barrett M, Ho K. *Potentially Preventable Hospitalizations—United States, 2004–2007*. Morbidity and Mortality Weekly Report no. 60(01) for the Centers for Disease Control and Prevention, January 14, 2011.

40. Newhouse J, Buntin MB, Chapman JD. "Risk Adjustment and Medicare: Taking a Closer Look." *Health Affairs*, vol. 16, no. 5, 1997.

41. Newhouse J, Huang J, Brand RJ, et al. "The Structure of Risk Adjustment for Private Plans in Medicare." *American Journal of Managed Care*, vol. 17, spec. no. 6, June 2011.

42. Newhouse J, Price M, Huang J, et al. "Steps to Reduce Favorable Risk Selection in Medicare Advantage Largely Succeeded, Boding Well for Health Insurance Exchanges." *Health Affairs*, vol. 31, no. 12, December 2012.

43. New York Times. "Medicare Bills Rise as Records Turn Electronic." Published online September 21, 2012. Accessed online at *www.nytimes.com/2012/09/22/business/medicare-billing-rises-at-hospitals-with-electronic-records.html?_r=0*

44. Park E. *Ensuring Effective Risk Adjustment: An Essential Step for the Success of the Health Insurance Exchanges and Market Reforms under the Affordable Care Act*. Washington, DC: Center on Budget and Policy Priorities, May 18, 2011.

45. Pope GC, Kautter J, Ingber M, et al. *Evaluation of the CMS-HCC Risk Adjustment Model: Final Report*. Report prepared for the Centers for Medicare and Medicaid Services, March 2011. Accessed online at *www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Downloads/Evaluation_Risk_Adj_Model_2011.pdf*

46. Pope GC, Kautter J, Ellis RP, et al. "Risk Adjustment of Medicare Capitation Payments Using the CMS-HCC Model." *Health Care Financing Review*, vol. 25, no. 4, 2004.

47. Schneider U, Ulrich V, Wille E. "Risk Adjustment Systems in Health Insurance Markets in the US, Germany, Netherlands and Switzerland." *CESifo DICE Report*, vol. 6, no. 3, 2008.

48. Shen Y, Ellis RP. "How Profitable Is Risk Selection? A Comparison of Four Risk Adjustment Models." *Health Economics*, vol. 11, no. 2, March 2002.

49. Thomas JW, Grazier KL, Ward K. "Economic Profiling of Primary Care Physicians: Consistency Among Risk-Adjusted Measures." *Health Services Research*, vol. 39, no. 4 pt. 1, August 2004.

50. United States Government Accountability Office. *Medicare: Per Capita Method Can be Used to Profile Physicians and Provide Feedback on Resource Use*. GAO-09-802 report to the Chairman, Subcommittee on Health, Committee on Ways and Means, House of Representatives, September 2009.

51. van de Ven WPMM, Schut FT. "Universal Mandatory Health Insurance in the Netherlands: A Model for the United States?" *Health Affairs*, vol. 27, no. 3, May 2008.

52. van de Ven WPMM, van Vliet RCJA, Lamers LM. "Health Adjusted Premium Subsidies in The Netherlands." *Health Affairs*, vol. 23, no. 3, May 2004.

53. Veazie PJ, Manning WG, Kane RL. "Improving Risk Adjustment for Medicare Capitated Reimbursement Using Nonlinear Models." *Medical Care*, vol. 41, no. 6, June 2003.

54. Winkelman R, Damler R. "Risk Adjustment in State Medicaid Programs." *Health Watch*, January 2008.

55. Winkelman R, Mehmud S. *A Comparative Analysis of Claims-based Tools for Health Risk Assessment*. Schaumburg, IL: Society of Actuaries, April 20, 2007.

56. Yu H, Dick AW. "Risk Adjusted Capitation Rates for Children: How Useful Are the Survey-Based Measures." *Health Services Research*, vol. 45, no. 6 pt. 2, December 2010.

Appendix II Interviews

To assess the areas of greatest need for improvement in risk adjusters, we conducted a small poll of stakeholders affected by risk adjusters, including representatives from CMS, state Medicaid programs, health plans, and model developers. Below, we summarize those findings and then draw our own conclusions from the literature and from these interviews.

Jean Yang is chief financial officer of Commonwealth Care, which serves as the provider of subsidized health insurance for Massachusetts individuals whose income is too low to afford the mandated coverage available in the private market or from an employer, but who do not qualify for Medicaid. Five private payers offer coverage with community-rated premiums. Payments are equalized with expected costs by transfers through the Massachusetts Connector, which collects premiums and distributes risk-adjusted payments using DxCGs. Because these plans are an extension of Medicaid managed care, data systems are available to collect information needed for DxCG algorithms.

Ms. Yang identifies two weaknesses of risk adjustment. One is that risk adjustment is never strong enough to overcome adverse or favorable selection. Because of this, providers must "adjust their business models" to stay in business. Plans participating in Commonwealth Care are public-spirited and will not maximize their selection opportunities. However, with the implementation of the ACA, new health plans will enter the market, and a wider range of beneficiaries will be served, changes which may increase incentives for selection and hence, the degree of selection that risk adjustment will confront. In addition, even if an effective risk adjuster is implemented, inadequate and inconsistent data is the norm in commercial markets and will be a challenge to overcome when ACA is implemented.

Daniel Zabinski is a senior analyst with MedPAC. He is one of several MedPAC analysts who has conducted risk-adjustment research to advise Congress on the MA program. This research serves as the basis for advice concerning risk-adjustment methods and calibration of payments with expected costs. He believes that current risk-adjustment techniques leave too much opportunity for health plans to benefit from favorable selection within risk categories. Two methods to overcome this problem are: allowing an additional year for accumulation of diagnostic information, as recommended by MedPAC, and incorporating additional illness severity measures in risk-adjustment algorithms. For Medicare, the most important severity adjustment would be for the likelihood of mortality, which is often accompanied by high expenditures.

Health Partners is a not-for-profit health insurer that is very active in Minnesota. We spoke to staff assembled by Sue Knudson, Vice President for Health Informatics. They described a wide range of risk-adjustment applications relevant to their activities, including rate-setting, underwriting and population health measurement. They described their underwriting model as containing several factors inappropriate for risk adjustment (for example, prior expenditures), and substantially more powerful than risk-adjustment models. This disparity in predictive power is a clear indicator of the potential for favorable or adverse selection. Besides this deficiency, they described several others characterizing most risk-adjustment applications. Perhaps the most significant is that risk adjustment as practiced provides little scope for health maintenance. Because risk adjustment in effect pays more when a patient's health deteriorates, and less when it improves, it provides little incentive for plans to invest in maintaining patient health. They also argued that one of the primary sources of favorable and adverse selection is propensity to consume or unmeasured need, and that risk adjustment cannot control adequately for that propensity. Because of this, plans offering generous benefits are subject to adverse selection,

and plans offering poor benefits experience favorable selection. For that reason, extra payments to generous plans and reductions in payments to plans offering limited coverage are needed to overcome biased selection.

Dr. Robert Master, the founder and CEO of Community Care Alliance (CCA) since the mid-1970s, identified important shortcomings of risk adjusters for both the Medicare and Medicaid programs. His insights are particularly valuable, because CCA, a network of providers dedicated to providing high-quality care for people with disabilities, has many dual eligible enrollees. Dr. Master feels that the HCC adjuster is much better than earlier risk adjusters and works reasonably well for most Medicare beneficiaries. However, his experience over many years is consistent with the empirical finding that the adjuster underpays for the individuals in the highest decile of predicted costs, while overpaying for those in the lowest decile. This failure is particularly critical for programs, such as CCA and a few others around the country, that seek to serve primarily those individuals at the high-cost end of the spectrum. He also notes that this failure is particularly bad for the under-65 disabled population, especially those with severe mental illness or substance abuse, some of whom are also homeless. Dr. Master feels that it is not possible to do the redesign of care delivery systems for this subpopulation without a specialized risk adjuster.

Despite these shortcomings of the Medicare risk adjuster, Dr. Master finds the bigger challenge for providers focusing on high-risk individuals to be the risk adjusters used by Medicaid programs. The risk cells defined are often too broad, and are therefore gameable and lead to underpayment for the highest-risk cases within a cell. For example, Massachusetts' Medicaid program defines four categories of Medicaid beneficiaries: those who don't need any long-term support services, those with cognitive impairments such as dementia, those who meet the criteria for nursing home entry, and those who reside in long-term-care facilities. While risk adjusters are used within these cells, they are not sufficiently refined. For example, individuals who were originally entitled to Medicare due to a disability and then turn 65 have a higher need for long-term support services than other elderly Medicaid members who meet the nursing home certifiability criteria. He suggests that the risk adjuster should include separate rates for these two groups, and for other subgroups of people who meet the nursing home criteria but have different levels of need for support services.

Dr. Master notes that the Minimum Data Set (MDS), collected on everyone in Massachusetts who meets the nursing home criteria and used in the risk adjusters to capture the type and level of support services needed, focuses primarily on physical disabilities, and too little on cognitive problems, which often are more important for determining need for support services. He suggests that a multiplicity of risk-adjustment systems are needed, with perhaps separate systems for plans that are delivery systems, such as theirs, versus plans that are insurers rather than providers.

Ultimately, Dr. Master believes that to adequately care for the highest-risk individuals and reduce their need for expensive services, it is necessary to adequately capture their true expected costs, so that the plan can afford the necessary investments to redesign care delivery.

Dr. David Knutson, a senior research fellow in the Division of Health Policy and Management of the School of Public Health at the University of Minnesota, has developed, tested, used, and written about risk adjusters for over 20 years, especially those used for Medicaid enrollees. Dr. Knutson noted that while a number of researchers have suggested adding functional status measures to risk-adjustment models, they actually yield little improvement in the predictive performance except for beneficiaries needing long-term support services. Furthermore, like diag-

noses, inclusion of functional status tends to result in upcoding of these impairments over time, so frequent recalibration would be necessary. Knutson also reiterated the concern over the models' perpetual underestimation of costs for high-risk cases, and suggests shifting to concurrent models. He noted that plans are actively conducting "predictive modeling" to find high-cost cases for intervention, and that these models outperform risk adjusters because they are not limited in what explanatory variables can be included in the model. In addition, some plans now have sophisticated electronic health records, which can provide data on lab tests and clinical factors such as ejection fraction that yield more valuable information on severity of illness than is available from diagnosis codes alone. Knutson argues that risk adjusters are even more important for capitating physicians or practices than health plans, and need to reflect quality as well as efficiency.

We also spoke with Dr. Richard Kronick, the Deputy Assistant Secretary for Health Policy in the Office of the Assistant Secretary for Planning and Evaluation at HHS, who before coming to HHS had developed the CDPS risk adjuster, initially designed for Medicaid beneficiaries. Like other respondents, Dr. Kronick cited as the key weakness of risk adjusters the reliance on data that are of poor quality, due to heterogeneity in the extent of upcoding of diagnoses. He suggests that other information be incorporated, such as multiple years of data or data from other sources, such as surveys.

In addition to these recommendations, we reviewed recommendations from the SOA's most recent comparative study (Winkelman and Mehmud, 2007). SOA's recommendations took the form of suggestions for further research rather than policy prescriptions. However, many of these suggestions dovetail with the recommendations described above. These recommended research topics included: the impact of plan switching behavior on the expected expenditures of the insured population, the robustness of models to different datasets, the predictive improvement resulting from longer data collection periods and from the inclusion of lab and other data sources, the development of predictive measures for different homogeneous population groups, and modeling for disease management uses of risk adjusters.

Appendix III Risk-adjustment applications

Risk-adjustment application	Benefit variability	Underwriting permitted	Choice of plan	Role of risk adjustment
Medicare Advantage	Minimum, may add at subsidized price, HMO, PPO, FFS permitted	No	Annual choice	Setting benchmark costs for health plans in relation to government option, compensating plans for variation in beneficiary selection due to benefit design, regional variations.
Medicare Part D	Actuarially equivalent, regulated formulary	No	Annual choice	Compensate plans for predictable variations in drug expenditures
Medicaid	Minimum required	Underwriting differs by state, usually varies by age and sex, set by competitive bidding, or regulated rates	Mandate, with default assignment, annual choice	Compensate plans for variations in expenditures not captured by rating cells. Selection may occur based on supplementary benefits, geographic variation, quality, network characteristics.
ACA Exchange	Benefit tiers, organization may vary	Rate bands, by age	Mandate enforced by tax	Equalize payments and costs across plans. Compensate plans for variation due to demographic factors, selection.
Swiss	Cost-sharing may vary	Age-rated premiums vary by region, income-related tax to risk pool	Mandate with annual choice	Equalize payments and costs across plans. Compensate for selection from benefit design, regional variations in cost.
German	Minimal, supplements may vary	Income-related premiums, risk-adjusted government contributions	Mandate with annual choice	Equalize payments and costs across plans. Compensate for variation based on selection through biased recruitment.
Dutch	May be FFS, capitated, benefits may vary, supplements may vary	Community-rated premiums, income-related tax to risk pool	Mandate with annual choice	Equalize payments and costs across plans. Compensate for variation based on selection from benefit design, income.

Benefit regulation limits the ability of health plans to induce favorable selection through their benefit design. If plans are limited in their ability to provide benefits appealing only to healthy beneficiaries, or place restrictions that beneficiaries needing treatment would reject, the extent of favorable or unfavorable selection is reduced. Thus the amount of variation in beneficiary risk for which risk adjustment must compensate is less.

Similarly, when beneficiaries have the ability to switch plans or choose among different plans, they are more likely to sort themselves by preference or need for health care services. Beneficiaries' freedom of choice also affords opportunities to payers to market selectively. When the menu of choices includes a government provider of last resort with limited control over consumption, incentives and opportunities for selection will be particularly strong. A market that places limits on choice, for example by assigning beneficiaries to plans, or placing limitations on their ability to switch, is one where the need for risk adjustment is lessened.

When underwriting is permitted, those factors incorporated in rate variation are no longer incentives for selection and need no longer be included in risk adjustment. However, when rates are set based on easily recognizable factors such as age, it does not matter on average whether payments to plans differ because of differences in premiums or risk-adjusted contributions from a health plan administrator.

The rightmost column describes functions fulfilled by risk adjustment. These functions may include both allocation of payments among plans, and determining payment rates. When used for allocation, risk adjustment is needed only to make comparisons between plans. When used to set rates, the formula must also determine the appropriate level of resource use. Risk corridors and reinsurance also reduce the amount of variation for which risk adjustment must compensate. Without these features, risk-adjustment models must also account for the extremes of variation.

Appendix IV Tests of changes to risk-adjustment models

Type of change	Base model and application	How change was tested	Result
Chronic conditions (16)	Medicare FFS, 1.1 million selected from 5% sample, 2004–2006, HCC model	Adding count of chronic conditions, additional chronic condition variables	Improved R-square, improved predictive ratio for chronic conditions. Under-prediction for the highest quintile of prior year expenditures.
Health status (22)	Sample from HMO population, regression based on survey comorbidities, 1990–1991	Adding results from SF-36 health status questionnaire, separate and interacted with other risk factors	With demographic variables and chronic conditions, an R-square of about 3% could be doubled by interacting chronic conditions with health status.
Health status (52)	Dutch sickness fund 1988–1994, 245,000 members, survey of 15,000 using DCG model	Adding survey measures of health status, separate and interacted with DCGs	R-square for a DCG model based on one year of data increased from 6.5% to 7.9% when survey data added. When DCGs were based on three years of experience, survey data increased R-square from 8.0% to 8.6%. Average losses for bad risks reduced.
Health status (56)	8,352 Medicaid children from Medical Expenditure Panel Survey (MEPS), 2000–2005, HCC model	Using a 2-part model, adding health status and children with special health care needs (CSHCN) survey information to HCC or prior use	CSHCN information added to HCCs increased R-square from 12.1% to 13.5%; predictive ratios for CSHCN increased from .55 to .97.
Prescription drugs (52)	2.8 million Dutch sickness fund enrollees, demographic model	Adding DCG diagnostic model and PCG drug model to demographic models	Average predictive losses in high-cost group reduced from 58% to 40% to 30%
Mortality (10)	Medicare FFS population, 5% sample of decedents, matched survivors 1994–1995, 150,000 total, PIP-DCG model	Modeling residuals from PIP-DCG payment model as a function of death and disease category	Coefficients on death large and significant
Extended history (16)	Medicare FFS, 1.1 million selected from 5% sample, 2004–2006, HCC model	Fitting HCC model on 2 years of diagnostic history	Predictive ratios for chronic conditions improved
Managed care data (41)	Selected Medicare Advantage plan 300,000 beneficiaries 2006–2007, HCC model	Comparing coefficients for HCC model estimated over MA to coefficients estimated by CMS over FFS	Impact of risk factors substantially different—R-square for managed care improved over cost factor estimated from FFS model
Truncation (55)	Commercial population, multiple models, 1998–1999, N=720,000, split halves testing; 2004–2005, N=720,000 split halves, various models	Truncating costs at $50,000, $100,000, $250,000	Improvement in predictive ratio for chronic conditions, high-cost cases. R-squares from the models from the 15%–22% range to the 21%–28% range when outlier costs are truncated at $100,000. When costs are truncated at $250,000, a lesser improvement in model fit and in predictive ratios at extreme values.
Previous expenditures (55)	Commercial population, multiple models, 2004–2005, N=720,000, split halves testing, various models	Adding first year expenditures	2 to 8 percentage point increases in R-square for prospective models, even when truncated
Concurrent modeling (55)	Commercial population, multiple models, 2004–2005, N=720,000, split halves testing, various models	Fitting various concurrent models with risk factors	Improvement in R-square ranging from 13 to 33 percentage points. Improvement in model fit for acute conditions, limited improvement for chronic conditions. Under-predicts substantially for those in the highest 5% of prior year expenditures.
Hybrid—concurrent model for selected conditions (13)	Commercial population, multiple models, 1998–1999, N=720,000, split halves testing, HCC model	Adding concurrent chronic conditions for the 9% of beneficiaries with chronic conditions selected by clinicians to prospective HCC models	Model based on prospective HCCs for most beneficiaries, but with concurrent HCCs resulted in a model R-squared of 36%, compared with an R-squared of 8% for a fully prospective model, and 37% for a fully concurrent model.

Appendix V Recent tests of risk-adjustment effectiveness

Study	Time period/ population	Method	Results	Implications
SOA (12)	Commercial populations 1998–1999	Commercial population, multiple models, 1998–1999, N=720,000, split halves testing	All models tested showed similar fit—substantial improvements to all from truncation, previous costs	Profits available from knowing previous cost, risk of high-cost cases promotes selection
SOA (55)	Commercial populations 2004–2005	Commercial population, multiple models, 2003–2004, N=720,000, split halves testing	Improvement in fit over 2002 study, model R-square, predictive ratio improved by truncation and including previous costs together or separately	Profits still available from knowing previous cost, risk of high-cost cases still promotes selection
Hsu et al. (23)	MA beneficiaries in a single MA-Part D plan, 2005–2006	Models based on RxHCC score, HCC risk factors compared with models of drug use or prior drug expenditures, fit over test dataset	R-square for models with prior spending fit best, models of drug use are next, both much better than RxHCC score	Opportunities for selection based on prior drug use in original Part D risk-adjustment model were quite large.
Frogner et al. (16)	Medicare 5% sample, 2004–2006	HCC model variants fit over sample, split halves test, chronic conditions, additional year for diagnoses	Predictive ratio based on number and type of chronic conditions improved, overall R-square remains ~10%, prior expenditures still predictive	Model fit can be improved for chronic conditions, profits from selection will remain
Pope et al., (45)	Medicare 5% sample, 2004–2007	HCC model variants fit over sample, split halves test	Predictive ratio for chronic conditions increased, overall R-square improves to ~12%	Revised HCC will improve fit for high-cost beneficiaries; profits from selection remain
Beck, Trottman and Zweifel (4)	Beneficiaries of one Swiss sickness fund from 1997 to 2004, N=180,000	Classify beneficiaries by risk type and simulate switching probability, measure expected profits and losses under different scenarios	Gains from attracting low risk, deterring high risk reduced by including prior hospitalization in risk-adjustment model, but minimally by pharmaceutical cost groups	Prior hospital use as risk adjuster reduces profits from selection, but profits from selection remain
Brown et al. (7)	Medicare Current Beneficiary Survey MCBS cost and use data 1994–2006	Compared prior cost for switchers into MA with predicted costs; compared MA enrollees before and after risk adjustment adopted, estimated likelihood of switching from MA to FFS as function of health status	Higher-risk patients more likely to enroll in MA with risk adjustment than without; lower prior expenditures within risk categories, lower spending compared with risk score associated with switching into MA, poor health associated with switching out	Expansion of covered population. Opportunities for selection remain; can be identified through prior costs; disenrollees from MA likely to be high-cost
McWilliams, Hsu and Newhouse (32)	Medicare MCBS sample, cost and use data 2001–2007	Compares risk score, health and resource use of MA new enrollees and those staying in FFS before HCCs, during phase-in, and after phase-in	Differences between enrollees and non-enrollees in risk, health, resource use, and risk of mortality all declined as HCCs were phased in	Risk adjustment expands the range of beneficiaries that can profitably be covered by MA.
Morrisey et al. (38)	Medicare FFS 5% sample 1999–2008	Fixed effects regression by county predicting enrollment, and prior costs of MA switchers relative to FFS stayers, and disenrollment from MA	Enrollment increased by HCCs, prior costs not significantly affected, number of disenrollees declined but costs of disenrollees increased	Risk adjustment expands coverable population, profits from selection reduced, favorable selection from disenrollment still powerful, may be prevented by cost-sharing for high risks.
Newhouse et al. (42)	Medicare FFS 20% sample 2003–2008	Compared risk scores and prior costs for switchers into MA relative to stayers, switchers out of MA relative to FFS, before during and after HCC implementation.	Risk scores of enrollees increased as HCC role in payments increased, prior costs controlling for risk did not change significantly. Post disenrollment costs of MA leavers increased when HCCs used.	Risk adjustment expands coverable population, profits from selection reduced, favorable selection from disenrollment still powerful, may be prevented by cost-sharing for high risks
MedPAC (36)	1.2 million beneficiaries in Medicare 5% sample 2007–2008	Compare risk-adjusted spending from 2007 of beneficiaries who switch into MA in 2008 with those who remain in FFS; compare risk-adjusted 2008 spending of beneficiaries who switch from MA to FFS in 2007 with those remaining in FFS.	Those who switched into MA in 2008 had 15% lower risk-adjusted 2007 costs than those who stayed in FFS; those who dropped MA had 16% higher costs in the following year than FFS who stayed FFS	Evidence for biased enrollment and disenrollment after risk adjustment; profit opportunities remain strong

For more information about the Synthesis Project, visit the Synthesis Project Web site at *www.policysynthesis.org*. For additional copies of Synthesis products, please go to the Project's Web site or send an e-mail message to *pubsrequest@rwjf.org.*

PROJECT CONTACTS

David C. Colby, Ph.D., the Robert Wood Johnson Foundation

Katherine Hempstead, Ph.D., the Robert Wood Johnson Foundation

Sarah Goodell, M.A., Synthesis Project

THE SYNTHESIS PROJECT

NEW INSIGHTS FROM RESEARCH RESULTS

RESEARCH SYNTHESIS REPORT NO. 25
JULY 2013

The Synthesis Project
The Robert Wood Johnson Foundation
Route 1 & College Road East
P.O. Box 2316
Princeton, NJ 08543-2316
E-Mail: synthesisproject@rwjf.org
Phone: 888-719-1909

www.policysynthesis.org