CHCS Center for
Health Care Strategies, Inc.

Predictive Modeling:
A Guide for State Medicaid Purchasers

August 2009

www.chcs.org

ACKNOWLEDGMENTS

Authors

Dave Knutson
Division of Health Policy and Management, University of Minnesota

Melanie Bella
Karen LLanos
Center for Health Care Strategies

Editor

Lorie Martin
Center for Health Care Strategies

This guide is a product of CHCS' *Rethinking Care Program*, which is developing and testing new strategies to improve health care quality and control spending for Medicaid's highest-need, highest-cost populations. It is made possible by Kaiser Permanente and the Aetna Foundation.

The **Center for Health Care Strategies (CHCS)** is a nonprofit policy resource center dedicated to improving health care quality for low-income children and adults, people with chronic illnesses and disabilities, frail elders, and racially and ethnically diverse populations experiencing disparities in care. CHCS works with state and federal agencies, health plans, and providers to develop innovative programs that better serve Medicaid beneficiaries. Its program priorities are: improving quality and reducing racial and ethnic disparities; integrating care for people with complex and special needs; and building Medicaid leadership and capacity.

For more information, visit **www.chcs.org**.

Table of Contents

I. Introduction

Medicaid purchasers are becoming increasingly interested in the potential value of predictive modeling (PM) to identify high-risk patients who are likely to benefit from care management interventions. While PM tools have historically been used to predict costs for rate-setting purposes, states can also use PM to identify "high-opportunity" candidates for care management and target public resources more effectively.

Predictive models are data-driven, decision-support tools that estimate an individual's future potential health care costs and/or opportunities for care management. Most commercially available PM tools classify individuals into future cost categories with a focus on high-cost cases. A few tools add a second component — commonly referred to as "impactability" — to identify patients who will potentially benefit from care management. Adapting PM tools to address the Medicaid population's intense and complex array of needs — ranging from physical and behavioral health comorbidities to socioeconomic issues — is a critical consideration for states that are planning to use PM. This holds true for states buying off-the-shelf tools as well as those with the analytical capabilities to build and/or customize PM tools in-house.

A predictive model can be described using three key features. The first is the *outcome* being predicted. Most PM tools include a model that predicts the relative future overall costs for an individual. Some PM tools also include models that predict other outcomes such as inpatient utilization. The second feature is the mix of *predictor variables* used in predicting the outcome. Predictor variables can include basic demographic information as well as diagnosis and prescription claims, functional status, prior utilization data, etc. Finally, a PM can be described by *how predictor variables are combined* to create the predicted outcome.

For predictive models to be sensitive to the complex needs of Medicaid beneficiaries with chronic illnesses and disabilities, each of the features described above should be considered. The predicted outcomes available should meet the specific needs of the population and the care management programs being considered. Further, the mix of predictor variables and the weights or rules assigned to each variable must reflect the population. For states interested in adopting PM, it is critical to ensure that the tool incorporate variables that are predictive for a Medicaid population. This is particularly important for states purchasing a tool, since many commercially available products were initially developed and/or validated on a non-Medicaid population.

One area of importance for Medicaid programs is behavioral health (i.e., both mental illness and substance abuse). Most PM tools include behavioral health predictor variables, however, few offer behavioral health predicted outcomes. A second critical area for Medicaid populations is psychosocial factors (e.g., housing

status, presence of informal caregivers, etc.) that are predictive of risk and/or the need for care management. Few PM tools address psychosocial factors in a comprehensive way. Obtaining these data and building them into the predictive algorithms can maximize the potential effectiveness of PM tools to identify individuals who might benefit from care management interventions.

To help state Medicaid agencies use PM tools to identify and prioritize candidates for care management, the Center for Health Care Strategies (CHCS) partnered with David Knutson from the University of Minnesota to develop *Predictive Modeling: A Guide for State Medicaid Purchasers*. This guide outlines key considerations for states to address prior to purchasing or building a PM tool. States interested in implementing PM can use it to:

- Understand which features of a predictive model are critical as well as how to enhance information that is derived from predictive models for Medicaid populations;
- Address planning questions to guide the implementation of predictive modeling; and
- Outline key considerations for choosing a PM tool to identify candidates for care management.

As states seek ways to more effectively target populations for care management, predictive modeling can be a valuable first step — yielding important information to help guide decision making. PM tools are best used in conjunction with other data to reveal as much information as possible to support identification and stratification of high-risk and potentially high-cost beneficiaries who are most likely to benefit from care management.

II. Guide to Predictive Modeling Tools

Most PM tools on the market today use customized data algorithms to predict individuals at risk of future utilization and/or costs. However, there can be major differences in the algorithms and breadth and depth of data used. Accordingly, states interested in PM for care management need a thorough understanding of the desired predicted outcomes, the data that can serve as predictor variables, and how these data combine to form a better picture of a person who may potentially benefit from an intervention.

Understanding Predicted Outcomes

Most PM tools offer users the ability to predict multiple outcomes. Clarifying the desired predicted outcomes that are being sought for a target population is an important step for states when using predictive modeling. The following list details the types of predicted outcomes that are available in PM tools.

Categories of Predicted Outcomes

- *Cost* – Almost all tools predict cost outcomes, usually relative costs. Costs could include total expenditures, costs related to type of service (e.g., inpatient, outpatient, professional, ancillary, pharmacy), or by condition. However, not many address condition-specific modeling.

- *Utilization* – The majority of PM tools predict utilization outcomes, usually relative use. Predicted utilization outcomes might include the likelihood of an inpatient admission or an emergency room visit.

- *Custom Cost or Utilization Outcomes* – States may need to customize a PM tool to calibrate cost or utilization outcomes to a particular population or outcome (e.g., a tool might be adjusted to predict long-term-care needs). When customization is required, a user could either handle in-house or seek the services of an external vendor.

- *Impactability* – Predicted impactability scores that link a relative value to the likelihood that an individual will benefit from care management are a critical function of PM tools. Impactability is a multi-dimensional assessment, incorporating the patient's perspective as well as the focus and capacity of a given intervention. The mix of conditions that are impactable, or actionable, for one care management program (e.g., serious mental illness, multiple physical chronic conditions, etc.) may not be a priority or capability of another program. For this reason, users should avoid off-the-shelf impactability scores that do not capture the multi-dimensional nature of this assessment. Purchasers of PM tools should ideally begin with the "end in mind" and start by identifying the desired

outcomes to help in clearly determining the data required to assess patient impactibility for the proposed intervention.

As an alternative to a pre-configured impactability score, tools that support rules-based algorithms allow users to model the patient population into cohorts, based on factors such as predicted risk, clinical profile, attitudes about health, healthy behaviors, and social context. Users can then assign a relative impactability score and return on investment value to each cohort taking into account factors that may influence the intervention, including a state's care management programs and capabilities.

Variables Impacting Predicted Outcomes

There are many variables that can influence predicted outcomes. Timing is a key factor that can cause variations. For example, a tool might identify future costs for an increment of time, typically ranging from three to 12 months.

Predicted outcomes can also vary significantly by population, depending on whether the population is commercial, Medicare, or Medicaid. Models for different populations can share a number of the same predictor variables, with the weights or rules used to combine the predictor variables differing depending on the outcome being predicted.

Glossary of Predictor Variables

Predictor variables are beneficiary characteristics that can be used alone or grouped with additional variables to predict an individual's future costs as well as the potential to benefit from an intervention(s). PM tools typically use multiple predictor variables to calculate a particular outcome. For example, enrollment and claims data (e.g., diagnoses, utilization, expenditures, etc.) are used to construct both the predictor variables and also the predicted outcome. Some models supplement basic claims data with additional data, such as functional status or information from health risk assessments or medical records, to further refine the targeting of cases or to identify cases before claims data are available. This enhancement is usually done after an initial run using a claims-based PM tool has narrowed the sample, thus making it more cost-effective to collect additional and often more resource-intensive and/or expensive data on a smaller set of beneficiaries.

This section details variables used in PM tools and provides information based on existing experience on the predictive accuracy of each variable.[1] Appendix A provides a brief review of the literature on the predictive contribution of different variables for PM.

---

[1] Information in this section is based on author Dave Knutson's experience in working with numerous states.

Assessing Data Variables for Predictive Models*		
VARIABLES	**PROS**	**CONS**
Age/Gender	▪ Easy to obtain ▪ Reliable and valid	▪ Poor predictive value, but useful when combined with other variables
Prescription Data	▪ Predicts costs almost as well as diagnoses (diagnosis proxy) ▪ Can be added to diagnoses to improve prediction ▪ Short claims lag	▪ Useful for care management planning, however potential for perverse incentive if used for provider or plan payment ▪ Requires frequent model updates
Diagnoses	▪ Best predictor of cost ▪ Required for care gap analysis	▪ Ambulatory coding inconsistent ▪ Long claims lag
Functional Status	▪ Critical for care management planning ▪ Useful for care gap analysis and intervention planning ▪ Can be collected upon enrollment	▪ Only moderately good predictive performance when used alone ▪ Not universally or reliably assessed ▪ Substantial additional costs for data collection
Utilization Data	▪ Important for identifying care gaps and potential cost savings ▪ Moderate improvement in predictive accuracy combined with diagnoses	▪ Useful for care management planning, however potential for perverse incentive if used for provider or plan payment
Prior Cost	▪ Good stand-alone predictor of future cost ▪ Relatively easy to obtain and use ▪ Adds modest improvement in prediction when combined with diagnoses	▪ Not as good a predictor as diagnostic, prescription, or utilization data ▪ Adds little clinical information ▪ Some costs, such as those around one-time medical events (e.g., a trauma-related hospital stay), are not predictive

* Based on author D. Knutson's work with states.

Claims Data

▪ *Diagnoses* – Diagnoses, an important predictor in any PM tool, provide good predictive power. They can also provide a useful context for adding other information, including diagnosis-specific utilization. In addition to enhancing accuracy, diagnoses also provide useful clinical information to understand the key drivers of a predicted outcome. Many PM tools assign a risk score to beneficiaries, with scores generally growing higher as the number and mix of significant diagnoses reported increases.

Diagnoses from some settings are more preferable than others. For example, compared to hospital-coded diagnoses, diagnoses from ambulatory settings are historically less reliably coded. PM tools should provide guidance on the

acceptable sources of diagnosis codes and have built-in checks that search for unusual or illogical patterns in the diagnosis data. A well-designed tool may exclude diagnoses from certain sources where coding may be tentative (e.g., ambulance) to reduce inaccurate data. A PM tool may also give little or no weight to those diagnoses that can be considered "rule-out," from services such as imaging and laboratory tests.

Uneven coding practices can be a mild to moderate problem for PM tools. For example, if a PM tool is implemented in an environment where reimbursement is linked to diagnosis-based risk adjustment, providers may be motivated to improve the specificity of diagnosis coding. These plans' members are likely to appear sicker than other plans that enroll beneficiaries with similar risks, but do not use diagnosis-based payment.

Claims inconsistencies over time in reporting major chronic illnesses is a particular coding problem that is likely to disproportionately affect PM tools for people with disabilities. For example, an analysis of Medicaid claims data from six states for people with disabilities found that over 43% of those coded with quadriplegia in year one did not have that code in year two. The lack of year-to-year coding continuity was found for many major chronic conditions. For multiple sclerosis, 42% of beneficiaries coded in the first year did not show up in the next year; for ischemic heart disease, 67% of beneficiaries did not appear in year two.[2] Since most PM tools use a 12-month timeframe for including diagnoses, if a diagnosis is not reported within one year, the PM tool assumes the condition does not exist.

Using a person with quadriplegia as an example, quadriplegia may be coded in one year, but not the next year — even if the patient received services — depending on how the provider(s) coded the reason for the visit. Thus, in this example, when the PM tool updates its 12-month window, it would miss the quadriplegia coding. While quadriplegia does not automatically indicate high medical costs, the high frequency of related acute conditions (e.g., urinary tract infections) would not get adequate value in the cost prediction model. This may be a missed opportunity, because the related acute care diagnoses could indicate good candidates for care management.

- *Prescription* – Prescription data, which are generally timely, reliable, and complete, are good proxies for diagnostic data — meaning that when used alone they are nearly as good as diagnoses at predicting future medical expenditures. Many prominent PM tools offer the option of using prescription data alone or in combination with diagnostic data. If diagnostic data are complete, adding prescription data achieves moderate improvement in predictive performance, particularly in cases where diagnostic information does not adequately reveal

---

[2] R. Kronick, T. Gilmer, T. Dreyfus, L. Lee. "Improving Health-Based Payment for Medicaid Beneficiaries: Chronic Disability Payment System." *Health Care Financing Review*. Spring 2000; 21(3):29-64.

condition severity. However, when diagnostic data are not complete, prescription data will often reveal cases that were not identified in the diagnosis record. For example, patients typically refill their prescriptions on a more regular basis than they visit their physicians. In addition, prescription data do not need to be obtained from providers, eliminating a difficult data collection step. Furthermore, prescription data can help identify care gaps, such as individuals who under-fill chronic medications.

While this guide is focused on PM for care management, it should be noted that the incentives for efficiency may be poor if prescribing is increased in order to raise a risk score. Prescription-based risk assessment models generally rely on drugs believed to be non-discretionary. However, to the extent that discretion remains in prescribing drugs for additional diseases or for less severe or marginal forms of the disease, caution should be exercised when prescription-based models are considered for payment applications.

- *Selected Procedures and Utilization* – Some PM tools look for prior selected utilization, such as an inpatient admission, emergency department (ED) use, or significant physician contacts, to refine high-cost case identification. The timing of the utilization can indicate importance, with more recent utilization events typically having greater significance. Providing a context for the utilization can also enhance predictive ability. For example, a recent inpatient stay for congestive heart failure carries greater weight in prediction than an inpatient stay for a lesser chronic or acute condition.

  In addition to risk prediction, procedure and utilization data are essential for PM tools that identify care gaps or perform some type of assessment of impactability (see care gap discussion on page 12). However, for states that are solely predicting future costs, these data add only a modest level of additional predictive value when added to a diagnosis-based PM tool.

- *Prior Cost* – An individual's total prior medical expenditures are a good stand-alone predictor of future expenditures. However, diagnosis-based PM tools predict future high-cost cases better than tools that rely on prior cost alone. Prior cost information added to a diagnosis-based model adds a low to medium level of improvement in future cost predictive performance. Further, since prior cost does not contribute any clinical information, prior costs related to some one-time medical events (e.g., a trauma-related hospital stay, delivered pregnancy) can lead to errors in prediction for some patients.

Clinical Data

- *Lab Results* – As clinical data become more readily available, their use in PM holds some promise for improving prediction and providing a more complete clinical picture for a patient. One source of clinical data that is becoming increasingly available is lab test results. Some PM tools include models that

allow the incorporation of lab results in addition to diagnoses and other variables. Not all lab results are useful in prediction, in particular those describing an acute or short-term condition status. Lab tests that are purely diagnostic in nature are also of limited use, especially where the confirmed diagnosis is also recorded on a medical claim for a patient. However, some lab tests describe an advanced level of severity for selected conditions and can contribute to prediction. Lab tests measuring organ function and cancer tumor markers are examples. Given the small number of patients with extreme values on these tests, the general contribution to overall predictive accuracy is small. However, for the patients with extreme lab result values, the advantage of the added information for predictive accuracy is significant. Lab results also provide value in identifying care opportunities and impactability.

Demographic Data

- *Age/Gender* – All PM tools use demographic risk factors, primarily age and gender. These variables are relatively easy to obtain, reliable, and valid. While this information has poor predictive value when used alone, these variables augment the predictive performance of a diagnosis-based model.

- *Socioeconomic Status/Living Arrangement* – When a PM tool is applied solely to a Medicaid population where most beneficiaries are in a similar economic bracket, economic status is generally not a useful predictive risk factor. This is in contrast to models used in commercially insured populations. Variations in social factors within a Medicaid population, however, may be useful in identifying candidates for care management or supports. Living alone, for example, is a risk factor that may or may not predict medical costs, depending on circumstances, but may be an indicator of the need for care management.

Survey Data

- *Functional and Health Status* – Functional and health status information from beneficiary surveys are recognized in many studies as moderately good predictors of future medical costs when used alone. In particular, this data source is often recognized as a potentially important predictor of medical expenditures for people with disabilities and is also an important indicator of the need for care management. However, these data provide a low level of additional predictive performance of future cost when added to diagnoses.

Functional status is not part of the routinely available enrollment and claims data and can be highly resource intensive to collect. Studies have also demonstrated significant respondent bias. Gaming is also a concern if the beneficiary or the proxy assumes that the answers will be used either to justify more services or take them away. States can consider using functional status or medical record data to enhance their PM tool; however, such enhancement is usually done after an initial screening using administrative data has narrowed

the sample, thus making it more cost-effective to collect additional and often more expensive data.

State Medicaid programs, health plans, or disease/care management entities often use screening surveys with new enrollees to identify the potential need for care management. In addition to a functional status component, these surveys may solicit self-reported information about health status, height and weight (i.e., to determine body mass index), risk behaviors, chronic diseases, and prior utilization. When used alone, the cost prediction performance of self-reported illnesses and utilization information is significantly less than claims-based diagnosis data. However, while somewhat unreliable, such self-reported data can augment targeting efforts for care management and/or facilitate early intervention until actual claims data are available.

Identifying Clinical Care Gaps and Care Opportunities

One common way to identify candidates for care management is by flagging "care gaps" between recommended care and what is actually provided to a beneficiary. Although care gaps are not typically used in the actual prediction of risk, they provide useful information to supplement measures of risk. A needs assessment is often the first step in determining where there are care gaps and who might benefit from an intervention. Tool developers and large health plans often assemble expert clinical panels to develop care gap logic that can be applied to claims data to identify beneficiaries who are not receiving recommended care and those with the greatest potential for cost savings. The logic, which is typically based on guidelines developed by organizations such as the National Committee for Quality Assurance,[3] the American Medical Association, or medical specialty societies, represents the evidence basis in the literature or expert consensus. PM tool developers often create further care gaps using internal expert consensus and evidence. The care gap "engines" included in PM tools will search for the occurrence of recommended treatments or tests related to specific conditions to determine compliance with prevention and chronic illness care guidelines. Some care gap logic identifies an obvious gap, while other care gap logic may suggest further investigation. For the Medicaid population, particularly critical care gaps to target include untreated substance use disorders or inadequate mental health care.

Much research has been conducted on admissions and ED visits for certain chronic conditions that could potentially be avoided with improved care management. These ambulatory care sensitive conditions are very costly and represent an important quality issue. Such conditions, which are recognized as often avoidable with appropriate prevention and early intervention, are an important target for PM tools. For example, patients observed with one of these conditions as well as a high predicted risk for a future inpatient stay would be good candidates for care

---

[3] For example, the Healthcare Effectiveness Data and Information Set (HEDIS), a nationally recognized measurement set developed by the National Committee for Quality Assurance, can be used to help develop care gap logic.

Common Predictive Modeling Myths

MYTH ▶ Each PM tool uses a very different set of core predictor variables.

FACT ▶ Most tools use the same set of claims data (e.g., diagnoses, prescription, utilization, prior cost), but vary in the other types of predictor variables used.

MYTH ▶ Since PM tools take various information sources into account, there is very little that the user needs to do other than putting in data and waiting for the tool to spit out a score.

FACT ▶ Purchasers need to spend sufficient time planning and answering key questions (i.e., How do you plan to use the tool? What is your target population? What information do you want to get from the tool?) before a PM tool is purchased. Once the tool is selected and is producing data, the state will need to validate the results before applying the knowledge to care management programs.

MYTH ▶ A beneficiary's risk score (i.e., a score assigned to a beneficiary by the PM tool related to likelihood of future high cost based on the number and type of risk factors) is the only type of information that is needed to target interventions appropriately.

FACT ▶ A risk score is a starting point for prioritizing/stratifying members, but not the complete answer. States should take into account all aspects of a beneficiary's health status, including clinical profile, gaps in care, and also consider information that is not available through traditional data sources (e.g., functional status, social context, and health behaviors and attitudes), in addition to their risk score. Continuous and targeted data mining is essential.

MYTH ▶ A PM tool will automatically pull in all of the clinical information that is needed. Since clinical information offers the most predictive value, there is no need to supplement this with other types of data.

FACT ▶ Understanding (and knowing about) changes in a member's social context, their willingness to engage, etc. is essential to selecting candidates for care management interventions and effectively tailoring those interventions to meet their specific needs.

MYTH ▶ Using care gap logic will help a state identify all of its target population.

FACT ▶ Care gap logic is based on existing evidence-based, nationally recognized guidelines and measures, which typically do not address the complex needs of Medicaid beneficiaries with disabilities and multiple chronic conditions. Care gap logic is valuable in understanding opportunities for patients and populations. However, a state's care management team needs to collect more detailed data, including predicted risk and clinical status, to support the case identification process for beneficiaries with complex conditions.

management. Beneficiaries who have been admitted frequently for potentially avoidable exacerbations of asthma or heart failure symptoms may also be high-opportunity candidates for disease or care management. However, a number of the beneficiaries experiencing these events will not have had prior inpatient or ED utilization. PM tools that take advantage of the entire breadth of predictor variables, rather than limiting to a subset (e.g., ambulatory care sensitive conditions) would be beneficial in such cases.

Care gap analyses are not always part of PM tools. Care gaps must be selected carefully so that those with weaker evidence or insufficient data do not increase false positives to an unacceptable level. When applying care gap logic to identify which clinically complex beneficiaries have the most potential to benefit from a care management intervention, it is important to consider the sources upon which the care guidelines are based. For complex care management, nationally recognized indicators (e.g., HEDIS measures, evidence-based practices) do not typically address comorbid conditions. Thus, care gap logic should be applied within PM tools in conjunction with the full breadth of information available regarding a beneficiary's health status, particularly comorbidities. Applying care gaps for beneficiaries with comorbid conditions will require input from practitioners in the program and will often include program-specific rules related to significant red flags not found in the evidence base.

Other Information to Support Care Management

Many PM tools provide added functionality that can be used to categorize patient information and derive further measures that can support the development of appropriate care management interventions. Such information can more completely describe the clinical status of a beneficiary, their likely receptivity to care management, and the social network available to support fruitful engagement. This information can include:

- **Clinical Profile** – What episodes of care, conditions, complications, and/or comorbidities are present? What is the history of medical and pharmacy services?

- **Health Behaviors and Attitudes** – What can be inferred from past behaviors around health? Areas to look at include compliance with prescribed care, healthy behaviors (e.g., smoking status, nutrition, exercise), and indicators of willingness to change.

- **Clinical Team** – Who are the physicians most responsible for a patient's primary care? What is the provider loyalty ratio (i.e., is there continuity of care)? What about specialty care? Identifying these physicians has value for provider engagement.

- **Social Context** – What is the home situation, e.g., are supports available to facilitate access to medical care? What additional supports are needed? What is the patient's primary language?

Information on health behaviors and attitudes, clinical team, and social context are not often included as variables in a predictive risk model. However, this information can provide significant value in stratifying beneficiaries for care management and tailoring interventions.

Determining Predictive Accuracy

Predictive accuracy of PM tools can be measured in a number of ways. For tools that predict future cost, one test of accuracy is how much of the variation in medical costs at an individual level can be explained by the PM tool. This can be assessed by an analysis of the correlation between the predicted outcome from a PM tool and the actual future cost. Another way to consider predictive accuracy is in terms of relative prediction error using the ratio of predicted cost to actual costs for the defined target populations. For PM tools that identify future high-cost cases, the typical method of evaluating predictive performance is to determine how accurate the model is in classifying individuals into one or more defined target groups (e.g., high-cost cases). For these purposes, accuracy can be defined as the proportion of actual high-cost cases that were not classified by the PM tool (sensitivity) and also the proportion that were classified as future high-cost, but did not incur high future costs (specificity). See Appendix B for a brief technical description of these evaluation methods.

Commercially available PM tools have been assessed by independent evaluators, primarily for determining payment rates for health plans or providers. The Society of Actuaries has sponsored three comparisons of select tools for cost prediction for commercially insured populations.[4] In addition, the Disease Management Association of America published a predictive modeling guide primarily directed toward disease management companies.[5] However, a systematic evaluation of these tools has not been conducted for many of their expanding uses, specifically for Medicaid purchasing. In assessing whether a patient is a good candidate for care management, validation must be based on an understanding of the specific care management intervention and who can most benefit from it. If the program is relatively new and the literature does not adequately address the issue of which patients have better outcomes, a self-sponsored program evaluation may be needed

---

[4] D. Dunn, A. Rosenblatt, D. Taira, et al. *A Comparative Analysis of Methods of Health Risk Assessment*. Society of Actuaries, December 1995; R. Cumming, D. Knutson, B. Cameron, B. Derrick. *A Comparative Analysis of Claims-Based Methods of Health Risk Assessment for Commercial Populations*. Society of Actuaries, May 2002; and R. Winkelman and S. Mehmud. *A Comparative Analysis of Claims-Based Tools for Health Risk Assessment*. Society of Actuaries, April 2007. **http://www.soa.org/files/pdf/risk-assessmentc.pdf**.

[5] M. Cousins, et al. *Predictive Modeling Buyers Guide*. Disease Management Association of America, 2007.

to refine the predictor variables that are used. See Appendix C for a brief review of the research on PM tool performance.

 Both types of predictive error — false positives and false negatives — can be reduced with more relevant and reliable data and also with modeling logic that is optimized for a specific program. However, typically the end user will trade-off which type of error to reduce by setting loose or tight population-targeting criteria, in effect changing the size of the targeted population. This process of adjusting case identification thresholds to reduce one type of classification error often forces a trade-off by increasing the other. For example, sensitivity (proportion of true cases that were not missed by the PM tool) can be increased by classifying or predicting which beneficiaries will be in the broader category of the top 20 percentile of cost rather than the top five percentile. However, if the program is actually designed for future high-cost beneficiaries, the number of false positive cases will also increase. Thus, increasing the sensitivity in a PM tool can result in more false positives, whereas higher specificity can lead to increased false negatives.

Balancing both types of errors is often a key consideration in the PM tool's internal design and some tools may not allow user-defined alternatives. An example is the PM tool requirement regarding the number of occurrences of a diagnosis in a claims history needed to trigger the diagnostic categories that serve as a risk factor in case identification. Many require only one occurrence of a diagnosis. Others may require two or more occurrences of a code in the assessment period. Some PM tools will add further predictive weight to a diagnosis based on the number of observed physician interactions or other utilization events specific to that diagnosis. The single occurrence specification will identify more true cases if, as is often reported, many major chronic diseases are not recorded in every contact with a provider. "Over-coding," however, is also common due to the use of tentative diagnoses that remain in the claims record even if the diagnosis is not later confirmed. By only requiring one code, a single occurrence of tentative diagnoses is more likely to increase the number of cases identified that do not actually have the disease. On the other hand, requiring that two or more codes be present will reduce the false positives, but simultaneously increase the proportion of missed cases. Recognizing single

Considerations for Medicaid Purchasers

This section outlined the types of predicted outcomes and predictor variables that can be found in PM tools, and discussed how tools can support care management decisions. Basic questions for a Medicaid program to consider related to these issues include:

- **What predicted outcomes does a PM tool offer?** Are Medicaid-specific outcomes available? Does the tool offer outcomes that meet the needs of my planned care management initiatives? Can the tool accommodate additional outcomes that are developed for my needs, using the experience of my population?

- **What predictor variables are used?** Can additional predictor variables be added, such as psychosocial factors, to enhance the prediction?

- **What additional information is used to identify care management candidates?** How are patients with actionable intervention opportunities identified? What information is available to provide a comprehensive picture of the patient to support more appropriate interventions? Examples of useful information to augment predictor variables include predicted risk, clinical profile, patient-reported data, gaps in care, and social context.

occurrences, but also giving added weight for disease-specific utilization, may provide some ability to address both sides of this issue.

PM tools often impose different rules for different applications. There are many examples of such trade-offs. Sometimes the trade-off is due to data coding errors as in the examples above. Other times it is due to the underlying inability to predict the future. Programs that have limited care management resources may prefer to minimize potential wasted effort by optimizing models to reduce false positives; while advocates of expanding services to all in need may push sensitivity so that no one potentially in need is missed by the model.

In summary, the optimal use of a PM tool and its results will likely depend on its targeted use and trading off the implications of missed opportunities against the effort involved in intervening with a patient who may not be well suited for a care management program.

III. State Considerations for Using Predictive Modeling

PM tools generally perform well for Medicaid populations with disabilities because of the prevalence of high-cost, relatively stable chronic conditions. This population also typically has lower enrollment turnover, which means that there is more history to assess. The shortcomings of predictive modeling for Medicaid programs relate more to the medical and psychosocial complexities of the population, which play a major role in determining candidates for care management programs. Risk factors that are common to Medicaid beneficiaries with complex needs, including behavioral health, substance abuse, and/or high levels of comorbidities, rely on difficult-to-obtain data that are typically not available from claims and, thus, are not usually a component of PM tools. However, the state, unlike a commercial health plan or employer purchaser, can augment the PM tool to identify beneficiaries with both medical and social complexity.

Most commonly used PM tools on the market perform similarly for predicting future cost at an individual beneficiary level. What differentiates them is how well the model performs for the user's specific application. Differentiating factors may include features that have been designed into the model, such as how diagnoses are manipulated. For example, the initial step in most models is assigning the many thousands of diagnoses into diagnostic groups. While many tools follow similar logic for this assignment, some tools differ substantially, and some even choose not to assign certain diagnoses that are considered trivial for cost prediction. The tools may also differ in the next step, which is related to how diagnostic groups are further aggregated and how rules and statistical tests are established for assigning priority to diagnostic groups. Additional important features to differentiate tools include how decision rules on case identification have been built into the tool, the number of user-defined options to customize the tool, and the flexibility of end-user reports.

Beginning with the "end in mind" and working backward from desired outcomes can help states in the planning and use of a PM tool whether purchased or internally developed. This requires being very clear about the objectives for using PM, its target population, and overall goals/expected outcomes of a Medicaid care management program. This is especially important for states buying off-the-shelf tools from vendors. Because many vendor tools were designed for commercial health plans/populations, these products may not represent the goals and interventions that a state is considering for its Medicaid population. The remainder of this section lays out key considerations for states in using PM tools for care management identification purposes.

1. Planned Use

States can use PM for a variety of applications, including case identification, forecasting program costs, and/or evaluating the program's ability to identify those

needing care management. Thus, clarifying the end goal for using a PM tool is a critical first step. For example, if the state's goal is to target the highest-cost beneficiaries, then a PM tool that allows the user to set the cohort threshold of cost at the top one percentile is an important consideration. If cost savings is a primary objective, then a user would want a tool with well-developed evidence-based logic to help identify interventions that might result in cost savings — though being careful not to set unrealistic savings expectations based on a tool's prediction.

Clarifying the decision-making process that PM will support (e.g., case identification, intensity of intervention, etc.) is also an important preliminary step. New state programs are often implemented in an environment with little room for making the wrong decisions. For example, if the PM tool refers too many cases that are not appropriate, valuable staff time and other resources may be wasted. This is a major problem, particularly for programs with a limited budget and/or under scrutiny. Being clear upfront about who will be served by the program will help states determine what types of PM predictive errors are more acceptable.

For states purchasing a PM tool, being as specific as possible regarding the tool's use will ensure that the state (or its vendor) will assess not only the tool's general attributes, but also whether the tool can be used off-the-shelf or whether it will require extensive customization.

2. Target Population

In thinking through the target population that the PM tool will identify, understanding how the core variables in the tool combine to form a more complete picture of a person in need of an intervention is critical. Additionally, states will need to know how to turn this information into defined cohorts of target populations for the model to identify. The state should be specific enough in its inclusion and exclusion criteria to identify the eligible population. If a key identification measure relies on data not available and/or not used in the typical PM tool, recognizing this early will help the state determine whether customization is necessary. Alternatively, the state may seek additional ways to identify the target population(s), such as a user-directed screening process. Finally, the state should decide whether the PM tool will identify cases that are referred to the program for further screening or whether the tool will trigger a referral or an intervention.

3. Expected Outcomes

Once the intended use and target population have been defined, then determining the information that is expected from the tool should be relatively straightforward. If the PM tool is used as the first selection criteria in a multi-stage screening process, it can produce a comprehensive profile of each case identified, determine whether a beneficiary is in the target group, and outline population distribution. Information about actual diseases of each "candidate" beneficiary can be used to support further screening as well as future intervention planning.

Building a Predictive Modeling Tool: Washington's Experience

In February 2009, Washington State's Department of Social and Health Services (DSHS) launched its internally developed predictive modeling tool.* Washington's tool, which is based on the Chronic Disability Payment and Medicaid-Rx Systems,** uses a variety of information to target care management needs, including: diagnoses; health risk drivers; drug use; preventive care opportunities; emergency room use; provider contact information; and mental health, substance abuse, and long-term care service use. The secure web-based tool is available for authorized DSHS staff and providers.

CHCS interviewed the tool's lead developer, David Mancuso, PhD, senior research supervisor for Washington State DSHS, to glean pointers for other states that are considering developing a predictive modeling tool.

Pros of Building Internally

- **Ability to Customize.** The state can modify the tool based on its care management priorities, design results and outputs to optimize use by care managers, and add features that are relevant to Medicaid.
- **Enhanced Staff Competencies**. The state's data staff thoroughly understands the Medicaid program as well as the implications of data for predictive modeling and care management decision-making. The resulting better connection between data staff and care managers further supports care management needs.
- **Potential to Control Costs**. Internal development may result in reduced long-term costs compared to buying a commercial product. Start-up costs, including equipment and software, and FTE costs may be higher initially, but ongoing and per client costs may decrease in the long term. Washington's PM tool is estimated to have a total annual cost that is roughly 25% lower than the annual cost of licensing a commercial product.

Cons of Building Internally

- **Lack of Internal Expertise.** Need to hire and retain staff — including software engineers, health economists, and statisticians — who can build web applications and calibrate predictive models.
- **Limited Data Infrastructure.** Extensive server support is required for building an in-house system that depends on multiple database and web application servers.
- **Access to Care Gap Knowledge**. Care gap analysis used commercially tends to be "black box." There is a limited availability of public domain care gap business rules or software, including HEDIS specifications and the Agency for Healthcare Research and Quality's Prevention Quality Indicators. Free software could be out-of-date.

Key Ingredients for Washington's Predictive Modeling Tool

- **Staff Funding.** Required 1.5 FTE to develop the PM tool and approximately 3.4 FTE for maintenance. FTE requirements could be lower depending on the amount and type of the data that is collected (e.g., Washington's PM tool collects long-term care, functional assessment, substance use, mental health data).
- **Experienced Staff.** Staff working on the PM tool were experienced with web-application development and risk-modeling and were eager to build on this experience.
- **Data Integration Infrastructure.** State was able to design and implement its own PM tool because it had 10 years worth of experience with building a data integration system.
- **Understanding of a Commercial PM Tool.** Washington state staff had experience using a commercial PM tool, which was helpful in building its own PM tool. As part of the development process, Washington conducted a head-to-head comparison with the commercial tool to assess the in-house tool's ability to accurately predict.

*Washington is initially using its predictive modeling tool to identify high-cost beneficiaries who have chronic health conditions complicated by mental health and substance abuse issues for its *Rethinking Care Program* pilot. For more information, visit www.chcs.org.

**R. Kronick, T. Gilmer, T. Dreyfus, and L. Lee. "Improving Health Based Payment for Medicaid Beneficiaries: CDPS." *Health Care Financing Review*, Spring 2000, 21(3):29-64. T. Gilmer, R. Kronick, P. Fishman, T. G. Ganiats, et al. "The Medicaid Rx Model: Pharmacy-Based Risk Adjustment for Public Programs." *Medical Care*, 2001; 39:1188-1202.

Many PM tools offer user-generated reports that provide multiple perspectives on the population(s) and on selected cases. Supplemental training/education (available from a vendor or an outside consultant) may be helpful to guide states in fully utilizing the information produced by a PM tool.

4. Data

Determining what data are needed (and available) to populate the PM tool is another key consideration. Since the majority of tools use claims data, the same data issues apply to most of the tools. Some tools, however, have more rigorous data inclusion or exclusion criteria than others. If a Medicaid program has claims and enrollment data, then it will likely have most of the data needed for a PM tool. It is important to keep in mind, however, that the information produced by PM tools can be further enhanced by including non-claims data such as functional status or other types of self-reported data (e.g., health risk assessments).

Timeliness of data should also be addressed in determining how to best detect changes in beneficiary status. There generally will be a claims run out lag and some processing time lag before new diagnosis or other risk factor data are available. Needs assessment screenings conducted by providers/care managers can provide more timely information to supplement PM output — recognizing that Medicaid agencies may not have ready access to this information. For more rapid response, the state may seek to purchase or build a PM tool that uses prescription data, which is generally available in less than a week. Prescription data can signal many major health problems before the full claims data-based model can be applied. PM tools can also be programmed to generate red flags to identify priority changes in a beneficiary's status that appear in the claims history. This timely information can help states work around the processing time lag and expedite care management decision-making.

5. Implementation

Capacity and expertise are needed to run the PM tool, produce reports, and, even more importantly, leverage the information generated by the tool. Some states may have the necessary resources and capacity in-house; others may need to obtain outside expertise/consultants. One factor to consider is the frequency of beneficiary assessment (e.g., monthly, quarterly, or every six months) as it impacts the staffing necessary to administer the tool. If a state does not have sufficient staff resources, it will need to consider obtaining outside assistance.

A team of analysts (whether in-house or outside experts) and end-users can be used to oversee the implementation and continuous improvement of the PM tool. A critical function of the team is to educate staff in using the tool's output for decision-making. An implementation team representing different departments and staff levels will be better able to generate the internal buy-in and technical resources necessary

to conduct population assessments, monitor data quality, and support tool improvements.

Customization of the tool may require a higher level of expertise. Again, some states may have the in-house analytical skills necessary to acquire the basic claims-based PM tool and enhance the tool for their own needs. Others may need to contract with an outside consultant for the customization. For states with the internal expertise, external PM experts could be enlisted, as needed, to provide valuable best practice information.

6. Validation

Creating a process to validate whether the PM tool is targeting the appropriate beneficiaries is important. The tool will not necessarily provide all the information needed to determine whether a person should receive the intervention. For example, the PM tool may need to be augmented with information generated through a screening process, often conducted by a provider using a survey. It may be useful to create an information feedback loop between the practitioners delivering the care management service and the state in order to improve the accuracy of the referral process. This feedback becomes most critical when the PM tool is used to support primary care-based care coordination. The primary care provider has valuable information on the needs and progress of the beneficiary that is richer than the data that can feasibly be included in a broad-based predictive model that uses administrative data. The provider could benefit from having utilization information on all the medical care provided to the beneficiary (something they typically do not receive) and the state can benefit from obtaining richer data on program recipients.

IV. State Considerations for Choosing a Predictive Modeling Tool

Once the state defines its goals and how the tool will be used (see Section III), the following considerations can help users choose to buy or build a specific tool.

1. Design and Reporting Logic

It is important to determine whether the tool's design and reporting logic fit the defined use. The number as well as the type of a beneficiary's conditions provides the most powerful predictors of future cost. Most tools attempt to address both the "additivity" of multiple conditions (i.e., estimating effects based on the sum of the number of conditions) as well as "interactivity" (i.e., estimating effects based on the interaction of multiple conditions that may be greater or lesser than the result of simply adding them). Users should be aware that PM models often vary in the emphasis that is put on these two elements and understand the resulting implications. In addition, although the comparative predictive accuracy of PM tools is similar overall, tools may differ at the sub-population or cohort level (e.g., someone with a specific condition or a specific combination of conditions). This is partly because PM tools may vary in the number of conditions that they incorporate. Most use almost all known diseases while some exclude minor, acute conditions under the assumption that these conditions are not relevant to risk selection, do not represent significant per capita costs, and/or their inclusion may produce a clinically needless proliferation of these codes.

Another difference among commonly used PM tools is the assignment of diseases to risk categories. This process may produce categories that are too heterogeneous for a specific disease. For example, if a care management program was developed for individuals with a relatively rare condition, a PM tool might categorize that condition with a group of unrelated conditions because it was too rare to have its own risk category. A prevalent disease such as diabetes, on the other hand, has its own category in most of these tools. The state will usually need to select populations based on the presence of specific diagnoses (e.g., using a HEDIS-based care gap definition) for a care management program.

2. Calibration Using State or Vendor Data

A state can choose to calibrate the tool directly on its Medicaid population or use the calibration offered by the vendor. It was once assumed that PM tools could only be valid if they were calibrated on the actual population to which the PM tool would be applied. Yet, experience has taught that imported relative weights on risk factors can be sufficiently valid and stable if they are estimated on a different population with similar characteristics, e.g., Medicaid TANF, Medicaid non-dual SSI, Medicare, or commercial plans with similar benefits.

Thus, if a state is relying on a vendor to calibrate the tool, the reference population should be similar to the application population (e.g., developed on an adult, non-dual, SSI-like population). A key consideration for states is to understand the reference population as well as the tool's calibration method. For the reference population, states will want to know: Is it Medicaid, commercial, Medicare? Are there continuous enrollment requirements? Is it based on total members or only on patients? How are outliers treated?

It is often preferable for states to calibrate the PM tool directly on the state's population. This requires both a sufficiently large population and adequate data. If the accuracy of the PM tool needs to be maximized for the highest risk segment of the population, small numbers can present a problem in developing stable risk weights. Data used for developing risk factor weightings requires the highest standard for completeness and consistency. For example, for duplicate claim records, duplications of diagnoses can be tolerated in the PM assessment, whereas duplications of charges could cause significant errors. In some cases, the robustness and quality of the data used by the developer of a PM tool can outweigh the benefits of developing a new model calibration that uses less than adequate data.

3. Frequency of Logic and Calibration Updates

Whether a state imports or calculates its own data, weights must be updated at regular intervals (every one to three years) to account for changes in practice patterns, coding changes, or significant changes in benefit design. Prescription-based models may need to be updated more frequently because the relationship between the prescriptions used in the model and medical expenditures may change rapidly (e.g., changes in prescribing patterns and the kinds of new drugs on the market). Users do not need to recalibrate these models with each new set of drugs — just when the drug concepts change or periodically to reflect new technology, indications, or practice.

4. Methods for Prioritizing Impactability

If a scoring system for prioritizing impactability is offered by the PM tool, it is important to determine how the rules and the scoring scale were developed. Key questions for states include: What is the source of medical evidence for the rule? Which rules are commonly used nationally (e.g., HEDIS)? How often are new rules added or updated? Are the priorities that the vendor used in developing the scores similar to those of the program? If there is not adequate research on which beneficiaries can most benefit from a specific care management program, how should the state develop its own criteria? Can the prioritization process be adjusted or augmented by the state to reduce potentially costly misclassifications?

Prioritizing is usually done for the care gaps identified. Most care gaps are condition-specific. The priority assessment crosses conditions; but the interventions that they point to will be primarily condition specific. And while it is true that, even for beneficiaries with comorbidity, care management will necessarily address specific

disease management problems, the interacting influence of multiple problems may alter the priorities in ways that cannot be predicted by the tools. This is particularly relevant where the tool does not include risk factors that are ultimately important in triage and care planning. Therefore, users should be cautious of off-the-shelf scoring tools as mentioned earlier in this guide. There may be some exceptions to the condition-specific focus of the tool's care gap analyses.

5. Data Requirements and Quality Monitoring

States should consider the tool's data requirements and what methods will be used to monitor data quality. It is important, for example, to determine the type of data quality checks that are incorporated into the tool. PM tools will often use benchmarked prevalence of diagnoses and also logic that corroborates internal data to assure that a member truly has the health condition or health event being measured by the variable. Corroboration may be as simple as assuring that women do not have a diagnosis of prostate disease or as complex as looking for an indicator of a treatment that should always be provided to a person with a coded diagnosis. PM tools employ many rules to mitigate data coding error. Diagnoses are aggregated into larger condition categories so that multiple codes can signal the occurrence of a condition. Most tools use diagnoses only from limited sources (physicians, nurse practitioners, and hospitals), and not from ambulances or free-standing diagnostic centers to avoid too many "rule out" codes. For some applications, some tools require more than one occurrence of diagnoses, or require that the data set used for the tool retain at least three or more coded diagnoses for each encounter, using combinations of diagnoses and strongly associated treatments to corroborate the condition or the severity level.

6. Time-Lag Specifications

States should be aware of the tool's specifications related to time lag between the date of change in the beneficiary's health status (e.g., newly diagnosed condition) and availability of the information for the PM tool. In general, the longer the time lag, the greater the loss of predictive power. It is more difficult to accurately predict the far distant than the proximate future. Many PM tools that predict future costs will require a minimum of six months of eligibility history in order to capture major chronic conditions and comorbidities; therefore, the model requires continuous enrollment of six months before risk is assessed. This requirement may present less of a problem for Medicare or for Medicaid beneficiaries with disabilities who have more stable long-term enrollment. It can, however, present a problem for Medicaid TANF and other programs where there is often high enrollment turnover and eligibility may terminate after weeks or months.

In addition to the PM tool's specifications about minimum eligibility history, users must add on the lag time required for claims to be processed as well as the time it takes to implement the PM assessment. To reduce the time lag, states can use a hybrid prescription data model until diagnoses are available, or can identify cases

early using other means such as an initial health assessment. In addition, the PM assessments that are based on claims data can be conducted monthly rather than a longer period (i.e., quarterly or every six months). Some users are building PM into payment systems and running nightly or weekly for selected patients. Note: If the user chooses not to follow a PM tool's minimum eligibility specifications, new calibration of the tool may be necessary.

7. Simulation of the PM Tool

A simulation of the PM tool is a crucial first step before going live. It is recommended that states obtain strategic consultation to do modeling for the defined cohorts of beneficiaries to be referred to care management. This will allow for a thorough in-house vetting of the data set being developed for the PM tool and will allow users to review actual reports produced from the state's own population. Simulating how the PM tool will be used to support or change a care management strategy can help states in understanding the key outputs and data that will be most useful. Preferably, the simulation would occur as part of the PM tool selection process. At a minimum, it should be part of implementation planning so that any necessary adjustments or adaptations can be made before launching the PM tool.

8. Multiple PM Applications

It may be that risk prediction is occurring for multiple and/or different uses in a given Medicaid agency (e.g., development of capitation rates vs. identification of high-cost cases, or one tool in managed care vs. another tool in fee-for-service). If other departments in the Medicaid program are using a particular method, it may be less confusing if the same approach is used. However, using different tools can be beneficial if each is best suited for its application and if internal expertise exists to cross-walk between reports on similar topics produced by different tools. If the PM tool will be used for multiple management purposes, then the use that most greatly distinguishes the performance of candidate risk assessment methods can be given highest priority.

9. Costs

The upfront and ongoing costs of a PM tool are important considerations for states to justify the investment. In some cases a tool may be ready for off-the-shelf use versus other cases where modifications are required. Thus, pricing can vary extensively based on a number of parameters, including whether a tool is in the public domain and how much implementation and consulting support a state requires. In calculating potential costs, it may also be helpful to consider the increased or decreased administrative costs beyond licensing fees. For example, some organizations may find that over time a PM tool pays for itself by improving internal efficiencies, thus decreasing administrative costs.

V. Conclusion

Whether states build a PM tool in-house or buy a commercial product from a vendor, it is critical to ensure that the benefits and limitations of these models are understood. The goal of this guide was to provide an overview to help states understand the range of considerations in choosing and implementing a PM tool.

As states move forward in more precisely targeting care management approaches to meet the needs of specific groups of beneficiaries, the value of strategically applied PM will become even more apparent. In particular, as the use of PM becomes more widespread in Medicaid, these tools may benefit from including certain variables (e.g., psycho-social, socio-cultural, etc.) that will make them even more relevant for a Medicaid population.

VI. Appendices

Appendix A: Brief Literature Review on the Predictive Contribution of Different Variables for Predictive Modeling

Numerous studies have evaluated predictive variables for their performance in forecasting future medical costs for individuals either alone or in combination with other variables. Studies focusing on predictive models for payment have been extensively conducted on commercially insured, Medicare, and Medicaid populations. Conversely, studies examining the use of predictive modeling for high-cost case identification have focused primarily, although not exclusively, on commercial applications. While it is difficult to generalize across all types of performance evaluation methods or across all populations in which the variables have been evaluated, a common pattern emerges in these studies regarding the general comparative predictive performance of specific risk variables in forecasting future medical costs for individuals.

One common performance measure that is often employed even when additional evaluation methods are also applied is the proportion of variation in total annual medical expenditures predicted at an individual member level by a predictive model. This proportion is often expressed using the R2 statistic.

It is important to note that the R2 statistic is not the only method or even the most preferred method for many applications of predictive modeling. For example, for classifying individuals into "future high cost case" and "not future high costs" categories, analyses based on sensitivity and specificity are often used (e.g., for evaluating diagnostic screening tests). Those evaluations have not been as ubiquitously applied in the long history of published studies of predictive performance. That is changing and studies using tests of a model's accuracy in classifying individuals are increasingly being published.

Therefore, the R2 statistic remains the most ubiquitous least common denominator measure in the literature as well as a meaningful performance indicator for predictive models. Following is a brief overview summarizing the most common findings of studies regarding the predictive performance of specific variables based on the R2 statistic. This is followed by a brief overview of selected findings related to sensitivity- and specificity-based analyses.

Administrative Data Variables

The research findings on the proportion of variation of cost predicted at an individual level by a variable or multiple variables are quite consistent. The literature is too expansive to cite for the purpose of this brief review. What is commonly reported across many studies is that age and gender predict only a small (R2 = 1-3%) proportion of the variation.   Diagnosis variables are almost never used

without including age and gender because age and gender are reliable and inexpensive data and can be added to diagnoses with little additional cost. Diagnosis models with age and gender predict about 10 times better than models that rely on age and gender alone (R2 =10-18%).

Two other administrative data variables — prior medical costs and prescription drug use — compete with diagnoses for predictive power. Using the prior medical costs of an individual to predict future medical costs performs nearly as well as diagnosis models. Prescription utilization data can serve as proxies for diagnoses and also add some additional severity information. Prescription variables alone have similar and sometimes slightly better performance than diagnoses alone.

Kuhlthau, et al. (2005) evaluated six risk adjustment methods, including two pharmacy-based and four diagnosis-based. The study compared the predictive accuracy of the methods for the Medicaid beneficiaries. The pharmacy- and diagnosis-based models had similar predictive accuracy.

Most administrative data-based predictive modeling tools offer models that combine variables. These usually perform better than any model based on one variable alone.

Zhao et al. (2005) reported that prediction models using both drug and diagnostic data best predicted future total health care costs (R2 = 16.8%) more effectively than models based on drug or diagnostic data alone (R2 = 11.6 and 14.6% respectively).

Predictive Performance of Administrative Data Variables Evaluated Using Sensitivity/Specificity-Based Analyses

Rosen et al. (2005) evaluated prospective risk-adjustment models using two diagnosis-based models, a prior-utilization model, and combined models (prior utilization and diagnosis). Diagnosis-based models performed better than the prior utilization models in identifying a subgroup of future high-cost individuals with high disease burden and chronic diseases appropriate for disease management. The combined models performed best.

Meenan et al. (2003) reported that two diagnosis-based models performed similarly and also preformed similarly to a prior cost-only model in discriminating future high-cost cases with prevalence targets of both the top 1% and 0.5% (AUC, 0.83-0.86).[6]

Billings et al. (2007) demonstrated that a number of Medicaid eligibility and claims data-based variables can be combined to predict future hospital admissions and costs effectively without the use of a commercially available tool.

Functional status and self-reported health status variables have also been evaluated alone and in combination with the administrative data based variables described

---

[6] Accuracy is measured by the area under the ROC curve (AUC). See Appendix B, page 29, for a description.

above to predict future medical costs. Functional status and self-reported health status with age and gender predicts about one-half or less as well as diagnosis models (R^2 = 4-6%). When added to diagnosis information, functional status may add modest additional predictive performance for certain populations (additional R^2 = 1%) (Hornbrook et al., 1996; Fowles, et al., 1996).

Maciejewski et al. (2009) compared functional status and diabetes-specific survey information with a diagnosis-based risk adjustment tool and found that the diagnosis tool produced twice the predictive performance as the functional status information. De Salvo et al. (2009) compared the predictive ability of a general single item self rating of health with: the Short Form-12 (SF-12), the Seattle Index of Comorbidity, and a diagnosis-based model. The general single item self rating model predicted the top quintile of expenditures as well as the SF-12 and the Seattle Index of Comorbidity, though not as well as the diagnosis model.

New modeling techniques, such as data mining and the use of time relationships among events, are exploiting as much as possible from claims data and even functional status data. The addition of more detailed clinical data is seen as the most promising advance in predicting future medical expenditures. Clinical data, primarily the results of diagnostic tests, are not broadly available. Such clinical data are increasingly becoming automated and with advances in electronic medical records will potentially be more readily available. These variables are believed to be the most promising new variables for improving the predictive performance of predictive models that forecast future cost.

Appendix B: Methods of Developing and Evaluating Predictive Models

Evaluations of accuracy of predictive modeling tools employ a variety of techniques. The following briefly describes the most common methods used.

Linear Regression Analysis

The linear regression model analyzes the relationship between the dependent variable and a set of independent or predictor variables. The equation predicts the dependent variable as a linear function of the risk factor(s). These risk factors are adjusted so that a measure of fit to a regression line is optimized. Much of the effort in model fitting and in model performance evaluation is focused on minimizing the size of the unexplained variation in the dependent measure, i.e. the residual. The accuracy of predictive models in predicting future cost is often expressed as an R^2 statistic. R^2 is the proportion of variability in a dependent variable that is estimated by a predictive modeling tool.

Sensitivity and Specificity Classification Models

When a predictive model predicts who will be in a high-cost group in the future, the concept of classification error can be applied. The proportion of false positive classifications determines the specificity of the classification model and the proportion of false negative classifications determines the sensitivity. Both sensitivity and specificity are often expressed in a single test to assess the accuracy of the model under conditions where the threshold for assigning cases (i.e., yes or no) varies. The single test is the area under a plot of sensitivity / specificity along a continuum of classification thresholds. The plot is called a receiver operating curve (ROC). For high-cost case identification, the model usually classifies who in the future will be in a high-cost group (e.g., upper 5%). The ROC is related to the fraction of **true positives** (TPR = true positive rate) vs. the fraction of **false positives** (FPR = false positive rate) along the curve as the threshold for assigning a person in the group changes (e.g., from upper 10% to upper 1%).

Accuracy is measured by the area under the ROC curve (AUC). An area of 1.00 represents a perfectly discriminating test; whereas, an area of .5 represents simple chance, such as a flip of a coin. A typically applied guide for evaluations of diagnostic tests in accurately identifying true cases is as follows:

- .90-1 = excellent
- .80-.90 = good
- .70-.80 = fair
- .60-.70 = poor
- .50-.60 = fail

Observed to Predicted Ratios

In addition to using individual enrollee-level statistical analysis to evaluate predictive performance, it is possible for the predictive modeling tool to perform well for one subpopulation but remain problematic for different subpopulations. This kind of prediction performance evaluation is often expressed as predictive ratios that compare predicted expenditures to actual expenditures for defined sub-populations. A ratio of 1:00 is the most accurate. Ratios above 1:00 indicate over-prediction and ratios below indicate underestimation.

Evaluations of predictive ratios are of different types, each of which examines the extent to which a method overestimates and underestimates costs for different sub-populations. One approach segments the population by actual medical expenditure levels (by quintiles or deciles) and evaluates the predictive ratio for each segment; another approach identifies sub-populations using specific chronic conditions or functional status to evaluate predictive ratios. Finally, some evaluations create simulated health plans or programs with varying proportions of high-risk members (risk selection) and evaluate the predictive accuracy for the simulated plans. In some

evaluation studies, actual enrollees in the program or plan have been studied directly.

Appendix C: Research on Predictive Modeling Tool Performance

The most commonly used predictive modeling tools use diagnoses in combination with age and gender. Some use prescription data alone. Either of these base tools may combine prior costs into the predictive model algorithm. All variables may also be combined. The base tools are, however, either based on diagnosis or prescription drug data. Therefore the literature comparing tool performance typically includes comparisons of the base models. It is the comparison of the base models reported here.

There have been numerous studies that tested the predictive performance of one or another of these tools, but because the populations and modeling methods were not uniform across the studies, a comparison across this vast literature back to the late 1980s would require more qualifications than can be addressed in this brief review. The Society of Actuaries (SOA) has sponsored three claims data-based comparative evaluations of these tools (1995, 2002, 2007). The most recent included not only the diagnosis-based tools, but also the prescription drug-based tools (Winkelman, et al., 2007). The findings of the most recent SOA study (Winkelman, et al., 2007) are reported below. In reviewing other studies and considering how methods and populations differ, other studies generally report similar findings to the SOA study.

Because the SOA study addressed predicting future medical costs by individual or population group, the related evaluations that address classifying which individuals will be in a designated high cost group in the future is not addressed. The Society of Actuaries 2007 study compared the predictive performance of numerous tools using multiple evaluation methods on the same large claims data set. The R^2 performance of these tools in their most optimized performance configuration in terms of outlier treatment and adding prior cost (which would be common in predictive models used to identify high-cost cases) was 20.5 to 26.5%. These findings are considered on the higher end of performance for diagnosis-based predictive models in predicting future total medical expenditures at an individual level due to the optimization described above. Three models that relied on prescription data alone were also evaluated. The R^2 statistic ranged from 26.3 to 27.1%. In addition, some models that combined variables from administrative data were included. One model that included diagnoses and total prior cost produced an R^2 of 29.1%.

Appendices Reference List

J. Billings and T. Mijanovich. "Improving the Management of Care for High-Cost Medicaid Patients," *Health Affairs*, 26, no. 6 (2007): 1643-1654.

K.B. DeSalvo, T.M. Jones, J. Peabody, J. McDonald, S. Fihn, V. Fan, J. He, and P. Muntner. "Health Care Expenditure Prediction with a Single Item, Self-Rated Health Measure," *Medical Care*, 47, no. 4 (2009): 440-447.

J.B. Fowles, J.P. Weiner, D. Knutson, E. Fowler, A.M. Tucker, and M. Ireland. "Taking Health Status into Account When Setting Capitation Rates: A Comparison of Risk-Adjustment Methods," *Journal of the American Medical Association*, 276, no. 16 (1996): 1316-1321.

M. Hornbrook and M. Goodman. "Chronic Disease, Functional Health Status, and Demographics: A Multi-Dimensional Approach to Risk Adjustment," *Health Services Research*, 3, no. 3 (1996): 283–307.

K. Kuhlthau, T.G. Ferris, R.B. Davis, J.M. Perrin, and L.I. Iezzoni. "Pharmacy- and Diagnosis-Based Risk Adjustment for Children with Medicaid," *Medical Care*, 43, no. 11 (2005): 1155-1159.

M.L. Maciejewski, C. Liu, S.D. Fihn. "Performance of Comorbidity, Risk Adjustment, and Functional Status Measures in Expenditure Prediction for Patients with Diabetes," *Diabetes Care*, 32, no. 1 (2009): 75-80.

R.T. Meenan, M.J. Goodman, P.A. Fishman, M.C. Hornbrook, M.C. O'Keeffe-Rosetti and D.J. Bachman. "Using Risk-Adjustment Models to Identify High-Cost Risks," *Medical Care*, 41, no. 11 (2003): 1301-1312.

CA. Powers, C.M. Meyer, M.C. Roebuck and B. Vaziri. "Predictive Modeling of Total Healthcare Costs Using Pharmacy Claims Data: A Comparison of Alternative Econometric Cost Modeling Techniques," *Medical Care*, 43, no. 11 (2005): 1065-1072.

A.K. Rosen, F. Wang, M.E. Montez, C.C. Rakovski, D.R. Berlowitz and J.C. Lucove. "Identifying Future High-Healthcare Users: Exploring the Value of Diagnostic and Prior Utilization Information," *Disease Management & Health Outcomes* 13, no. 2 (2005): 117-127.

R. Winkelman and S. Mehmud. *A Comparative Analysis of Claims-Based Tools for Health Risk Assessment*, Society of Actuaries, April 2007. Available at **http://www.soa.org/files/pdf/risk-assessmentc.pdf**.

Y. Zhao, A.S. Ash, R.P Ellis, J.Z. Ayanian, G.C. Pope, B. Bowen and L. Weyuker. "Predicting Pharmacy Costs and Other Medical Costs Using Diagnoses and Drug Claims," *Medical Care*, 43, no. 1 (2005): 34-43.

Rethinking Care Program – Additional Resources

This guide is one of a number of tools being produced by the Center for Health Care Strategies (CHCS) through the *Rethinking Care Program.* This national initiative was developed by CHCS to serve as a Medicaid "learning laboratory" to design and test better approaches to care for the program's highest-need, highest-cost beneficiaries. The initiative is linking state pilot demonstrations — currently underway in Colorado, Pennsylvania, New York, and Washington — with a national learning network committed to advancing Medicaid's capacity to serve these "high-opportunity" beneficiaries.

For more information about the *Rethinking Care Program*, as well as tools for improving care management for Medicaid beneficiaries with complex needs, visit **www.chcs.org**.