CHCS

Center for
Health Care Strategies, Inc.

Resource Paper

Strategies for Assessing Health Plan Performance on Chronic Diseases: Selecting Performance Indicators and Applying Health-Based Risk Adjustment

By Ann Volpel
John O'Brien
Center for Health Program Development and Management
University of Maryland,
Baltimore County

Jonathan Weiner
Health Services Research and Development Center, Johns Hopkins Bloomberg School of Public Health
Johns Hopkins University

Funded by the Center for Health Care Strategies, Inc. under The Robert Wood Johnson Foundation's Medicaid Managed Care Program

March 2005

270

Table of Contents

Appendices*

* Appendices are available on the CHCS Web site.

Acknowledgements

The authors would like to thank Tricia Roddy and John Folkemer, Maryland Department of Health and Mental Hygiene, for their participation and guidance. We would like to thank the Advisory Committee members (listed in Appendix I) for taking the time to provide valuable feedback throughout the duration of the project. We are also grateful for the editorial assistance offered by Karen Llanos and Michael Ashworth from the Center for Health Care Strategies.

Finally, the authors would like to acknowledge the efforts of the entire project team, whose contributions were critical to its success: David Idala, Steven Johnson, Michael Nolin, John Kaelin, and Chuck Milligan, all presently or formerly employed by the Center for Health Program Development and Management, University of Maryland, Baltimore County; Hoon Byun, Health Services Research and Development Center, Johns Hopkins Bloomberg School Of Public Health, Johns Hopkins University; and Alice Burton, during her tenure at the Maryland Department of Health and Mental Hygiene.

Executive Summary

As Medicaid managed care programs mature, states are looking to refine their methods of measuring and improving the performance of participating health plans. Along with other private and public health insurers, states are developing strategies to "pay for performance" and implement "value-based purchasing." Many states are also shifting their focus to evaluate the care provided to enrollees with chronic diseases, as SSI eligibility groups are incorporated into managed care programs.

The main goal of our project was to pilot an approach to measure health plan performance using administrative data that other states could adopt to evaluate the care provided to enrollees with chronic diseases. There are many benefits to developing a reliable source of administrative data for use in performance measurement. States can assess the care delivered to specific populations (e.g., adults with disabilities). If a state is interested in evaluating performance in an area that is not addressed by existing HEDIS® measures, the state may develop its own targeted measures. The analysis of administrative data provides states with the ability to examine data from several perspectives and "drill down" on unanticipated results. Overall, the use of administrative data provides states more flexibility in analysis.

Developing a performance measurement program requires states to consider a number of factors. The process can be divided into five steps. Our analysis of this process produced the following lessons for states:

- *Selecting the Data.* States must understand the strengths and limitations of the data being used to measure health plan performance. Most states will select some type of administrative data set for the analysis. The quality of encounter data submitted by health plans may vary significantly depending upon the incentives for submitting complete and accurate data and the nature of the contracting relationships with providers. Understanding the data limitations that affect the accuracy of performance results is critical to establishing an effective program.

- *Selecting Diseases for Analysis*. Establishing a disease-specific performance measurement program requires a state to identify a set of diseases for analysis. There are a number of factors to consider when selecting diseases. The individual goals of each state will influence the priority it places on each factor. The following set of decision rules should be considered when making the selection:

  The incidence of the disease should be reasonably prevalent across the health plans. A large population size will not only increase the validity of the analysis, but also increase the likelihood that the health plans will take an interest in the results.

  The disease must have clinical standards of care and those standards should include process measures that are strong predictors of positive outcomes. A successful

performance measurement program will evaluate health plan performance against a well-defined and accepted standard.

The performance of the health plans in treating the disease must be measurable with available data.

The potential for quality improvement and cost savings within the Medicaid program should exist. The feedback provided through a performance measurement program should encourage health plans to improve performance, with the additional goal of improving health care outcomes and providing cost savings, where appropriate.

- *Defining Disease Cohorts.* Specificity in defining the cohort is more important than the size of the cohort when monitoring performance on individual diseases. States must establish both enrollment and clinical criteria when defining a cohort for analysis. The parameters selected for each set of criteria will affect what a state is able to measure. For the decisions about enrollment criteria, the trade off is between the size of the population being assessed and continuity of enrollment within a single health plan. On the other hand, decisions about clinical criteria require a trade off between the size of the cohort and a guarantee that all members of the cohort have the disease condition being studied.

- *Selecting Performance Measures.* Once a state has decided which diseases to include in the study and defined the cohort of individuals with each disease, the next step is to select performance indicators on which to evaluate health plans. Performance indicators:

  Should be measurable with available data. States should test the data to ensure that the information being measured is available and reliable.

  Should be based on nationally recognized benchmarks, standards of care, or research. By applying this second criterion, a state will increase the likelihood of acceptance by the health plans and their providers.

- *Involving the Health Plans.* When designing a performance measurement program that will evaluate the performance of specific health plans, it is important to gain acceptance from the health plans that will be evaluated. States should consider including health plans in the planning process to find out which diseases they have been focusing on and whether they have concerns about the appropriateness of the measures being considered.

The second goal of our project was to identify performance measures that are associated with desired outcomes and would be useful to include when evaluating health plan performance in any state's Medicaid managed care program. In general, administrative data are a poor source of information for measuring medical outcomes. Administrative

data do not contain the information needed to answer whether an individual's health status has improved over time. However, some utilization measures can serve as useful indicators of positive medical outcomes. Inpatient admissions are the best example, as a general consensus exists that lower admission rates are an indicator of "better" care. Our analysis explores service utilization and demographic factors that are associated with improved medical outcomes for each disease we studied. This analysis provides a foundation for the selection of indicators that a state may consider including in a performance measurement program.

A unique feature of the approach presented in this report is the application of diagnosis-based risk adjustment to performance measurement. The application of diagnosis-based risk adjustment to performance measurement within Medicaid is a relatively new field; it is more commonly applied to payment mechanisms. Risk adjustment offers great value for performance measurement as it provides the ability to reduce the potentially confounding effects of case mix differences between health plans.

Whether or not to include risk adjustment in a performance measurement program is not an insignificant decision for states. States that have the capability of adjusting for risk must first decide which measures are appropriate for adjustment. Our analysis suggests that it is useful to risk adjust when:

- *The measure varies significantly across the risk factor of interest.* If the rate of hospital admissions is the same for all enrollees regardless of their relative health status, then risk adjustment will have little to no effect on performance results.
  The risk factors vary by unit of analysis (in most cases, health plans). If the case mix of each health plan is the same, then there is no reason to adjust the results.
- *There is a conceptual argument supporting the need to identify the "independent" impact of the health plan on care.* If the standards of care suggest that all individuals with a disease should have a certain test, it would be difficult for a state or a health plan to argue that the performance results should be adjusted for health status, regardless of the level of variation.

States that pursue risk-adjusted performance assessment must also decide which factors are appropriate to include in their risk adjustment. Demographic factors such as age, sex, region, and eligibility category are most frequently included. States may also include health status. By including a variable in a risk adjustment, a state is implicitly conveying that any variation in performance across health plans because of that factor is acceptable and that the plans are not expected to overcome that variation through outreach, case management, or other means. In general, states should adjust for only those risk factors that are independent of health plan control and that might have a direct impact on plan performance.

Following the decision rules described above, we developed a performance measurement program using Maryland's Medicaid data to analyze the care provided to enrollees with asthma, diabetes, HIV/AIDS, and schizophrenia. Using a combination of process and

outcome measures, we analyzed the performance of the six health plans participating in the Medicaid program in CY02. The results are risk-adjusted, where appropriate. Our analysis produced the following results:

- *There is a strong direct relationship between health status and utilization rates for inpatient admissions and emergency room (ER) visits.* This suggests that health plans with a sicker case mix would be expected to have a higher percentage of enrollees with inpatient admissions or ER visits. Applying health-based risk adjustment to the results for these indicators improves the accuracy of the measurement by controlling for any variations in case mix across the plans.

- *Performance on process measures is not generally sensitive to health status.* These results are consistent with clinical expectations. Standards of care that are appropriate for all enrollees with a certain diagnosis should be applied to all enrollees evenly regardless of health status. Applying health-based risk adjustment is not suitable for such indicators.

- *Receiving ambulatory care services is associated with a decrease in the likelihood of an inpatient admission.* When controlling for other factors (including health status), enrollees who had two or more ambulatory care visits were approximately one-third less likely to have an inpatient admission than those enrollees who had fewer than two ambulatory care visits. This conclusion is consistent with the literature on ambulatory care sensitive conditions, which suggests that appropriate outpatient care can reduce the need for inpatient admissions for certain health conditions. Most states and health plans would agree on a goal of reducing inpatient admissions. Hence, this analysis further supports evidence that ambulatory care visits can decrease an enrollee's likelihood of admission, thus supporting the inclusion of an ambulatory care visit threshold in any state's performance measurement program. Even states that do not have sophisticated data systems to stratify enrollees by morbidity or to risk adjust performance results can likely document ambulatory visits.

The information contained in this report is useful for any state that is interested in evaluating health plan performance using administrative data, regardless of whether it has the capacity to add diagnosis-based risk adjustment to the analysis. States with well-developed managed care programs will benefit from the more detailed presentation of the health plan profiles.

Introduction and Background

As Medicaid managed care programs mature, states are looking to refine their methods of measuring and improving the performance of participating health plans. Along with other private and public health insurers, states are developing strategies to "pay for performance" and implement "value-based purchasing." In advancing these goals, states face many decisions: What should be measured? What threshold defines "good" care? Against what standard shall health plan performance be measured? The process of answering these questions and establishing a performance measurement program requires a significant investment of time and resources for a state.

To measure the performance of participating health plans, many Medicaid programs rely on health care quality indicators such as the Health Plan Employer Data and Information Set (HEDIS®). Using HEDIS® measures offers states many benefits. It is a standard set of metrics by which to evaluate health plans: it provides an opportunity for multi-state and commercial comparisons, the work of defining and testing indicators is performed by a reputable outside organization, data analysis is the responsibility of the health plans, and the results are audited by certified organizations.

HEDIS® measures do have their limitations, however. The majority of HEDIS® measures do not address care to enrollees with chronic diseases. Chronic diseases are especially important as states consider including Supplemental Security Income (SSI) eligibility groups in managed care.[1] Health plan performance is measured only annually and the data requirements for some indicators is extensive (e.g., clinical lab values, provider specialty, etc.)

This project tests a supplemental method of measuring health plan performance exclusively through the use of administrative (encounter) data. There are many benefits to developing a reliable source of administrative data for use in performance measurement. States can assess the care delivered to specific populations (e.g., adults with disabilities). If a state is interested in evaluating performance in an area that is not addressed by existing HEDIS® measures, the state may develop its own targeted indicators. The analysis of administrative data provides states with the ability to examine data from several perspectives and "drill down" on unanticipated results. States may also combine an analysis of service utilization in the managed care program with an analysis of any carved out services. Furthermore, a state may evaluate the data on a more frequent basis (e.g., semi-annually instead of annually). Overall, the use of administrative data provides states more flexibility in analysis.

The main goal of our project was to pilot an approach to measure health plan performance using administrative data that other states could adopt to evaluate the care provided to enrollees with chronic diseases. A unique feature of our approach is the

[1] *Maryland has included segments of the SSI eligible population in its managed care program since its inception.*

application of health-based risk adjustment. The application of health-based risk adjustment to performance measurement within Medicaid is a relatively new field. It offers great value for performance measurement as it provides the ability to reduce the potentially confounding effects of case mix differences between health plans. Using encounter data from Maryland's Medicaid program, we apply risk adjustment to our measurement of care provided to enrollees with four chronic diseases: asthma, diabetes, HIV/AIDS, and schizophrenia.

The second goal of our project was to identify performance measures that are associated with desired outcomes and would be useful to include when evaluating health plan performance in any state's Medicaid managed care program. Administrative data are generally perceived to be a poor source of information for measuring medical outcomes because they do not contain the information needed to answer whether an individual's health status has improved over time. However, some utilization measures can serve as useful indicators of positive medical outcomes. Inpatient admissions are the best example, as a general consensus exists that lower admission rates are an indicator of "better" care. Our analysis explores service utilization and demographic factors that are associated with improved medical outcomes for each disease we studied. This analysis provides a foundation for the selection of indicators that a state may consider including in a performance measurement program.

This report is designed to assist states in developing their own performance measurement programs using administrative data to evaluate the care provided to health plan enrollees with chronic diseases. Section I addresses the factors to consider when developing a program and the criteria that may affect each decision rule. It also contains a summary of recommended performance measures based on empirical research. Section II explores the application of diagnosis-based risk adjustment to performance measurement. We discuss the risk factors to include in the adjustment process and demonstrate how risk adjustment can be effective. We also describe the factors that influence when the application of risk adjustment is appropriate. In Appendix II, we apply the "lessons learned" in Sections I and II to develop a profile of the performance of six Medicaid health plans from Maryland on treating enrollees with four chronic diseases. Evidence to support the selection of performance indicators that are associated with improved medical outcomes is also included.

The research and analysis was conducted by the Center for Health Program Development and Management (Center) at the University of Maryland, Baltimore County (UMBC), with support from experts in risk adjustment and medical practice from Johns Hopkins University Bloomberg School of Public Health. The Maryland Department of Health and Mental Hygiene (Department) provided consultation throughout the project.

The information contained in this report is useful for any state that is interested in evaluating health plan performance using administrative data, regardless of whether it has the capacity to add diagnosis-based risk adjustment to the analysis. States with well-

developed managed care programs will benefit from the more detailed presentation of the health plan profiles.

Section I: Decision Rules for Developing a Disease-Specific Performance Measurement Program

Any state that is interested in developing a disease-specific performance measurement program must evaluate a number of factors. What data are available? Which diseases should be studied? On what measures should health plan performance be evaluated? In Section I, we address the major decisions that each state must make to develop its own program. We also provide an overview of the factors that may impact each decision. Finally, we provide examples of how our project team evaluated each factor and made its decisions.

Data

Several different types of data are used to create performance indicators. If the available data are of unacceptably poor quality, performance measurement may be impossible. Fortunately, the creation of performance measures is not an all-or-nothing exercise. Valid, useful performance measures can be created even when some data are limited in regard to reliability or validity. A state's understanding of the strengths and limitations of its data, however, is critical to developing an effective measurement program.

Three types of data are combined to create performance measures: eligibility and enrollment, diagnosis, and service.

- Eligibility and Enrollment Data. Eligibility and enrollment information is fundamental to program management and payment. Maintained centrally, this information is generally high quality and reliable. Eligibility and enrollment data are used to select individuals who meet enrollment criteria (such as those required by HEDIS®). These data also include demographic information such as age, sex, and eligibility group (SSI vs. Temporary Assistance to Needy Families (TANF)).

- Diagnosis Data. Physician and hospital claims are rich sources of diagnostic information that can be used to identify individuals with specific conditions. Claims, however, are just one way to identify and track individuals with specific conditions. Alternative methods, such as registries that can be linked to eligibility data, or special payment categories (e.g., HIV/AIDS, chronic conditions), can be equally reliable sources for identifying individuals with specific diagnoses.

- Service Data. Once individuals with a specific condition and eligibility period are identified, the process of performance measurement begins to examine whether certain services occurred. Most Medicaid managed care programs require health

plans to submit encounter data for the services provided (although some may only request a limited data set). Encounters fall into several broad categories; the relative strengths of each type of data will influence the type of performance measures on which a state may rely.

- *Physician Services.* The extensive use of fee-for-service contracts among managed care organizations means that physician data are often of good quality and complete. Health plans that make extensive use of sub-capitated contracts (whereby physicians are paid a monthly fee) tend to have more difficulty submitting complete physician data.

- *Hospital Services (inpatient and outpatient).* Hospital services, particularly admissions, are costly events that health plans will carefully track within their internal data systems. In addition, hospitals tend to have sophisticated billing systems (relative to private physician's offices) that should lead to high quality data. Limiting the potential of these data is the complex nature of hospital claims (e.g., a large number of separate services or elements that may be initially rejected and resubmitted). The complexity of the claim may lead to data submission challenges that limit the completeness of inpatient services data.

- *Pharmacy Data.* Electronic submission of pharmacy data has long been the industry standard. Pharmacy claims, therefore, tend to be very complete and reliable.

- *Other Service Types.* Lab tests, x-rays, and other services can be key elements in various performance measures—the provision of lab services being especially useful. The quality of these data is affected by health plan contracting practices, in that health plans may subcontract a bundle of services (e.g., lab, DME) at a flat rate. In Maryland, lab data for hospitals are all billed under a single code, making it impossible to distinguish which lab tests were provided. In these circumstances, the utility of the data submitted to health plans, and ultimately the state, is often variable.

The primary data source for this project was health plan encounter data. Maryland benefits from a rich source of encounter data in its Medicaid managed care program. Significant efforts by both the Department and the participating health plans have continued to improve the collection and processing of encounter data since the start of the managed care program in 1997. Maryland is now considered a national leader in encounter data collection as a result of direct technical assistance to health plans, development and dissemination of useful and timely feedback reports, and population-based projections of encounter data volume. The strength of Maryland's data is demonstrated by the state's willingness and ability to use the data for essential program functions, such as risk-adjusted payment rates; calculating most key indicators used in a five-year evaluation of the program; and calculating specific measures for Maryland's

value-based purchasing program. Furthermore, and perhaps most important, the health plans have an incentive to submit timely and accurate encounter data because the diagnostic information is used to risk adjust the capitation payments to the health plans. Although the current managed care program in Maryland carves out mental health services, mental health claims data were also included in our analysis for selected diseases.

The Center has developed a variety of methods to test the completeness of the data it collects to more easily identify gaps or missing data that may affect the data's application. We have relied on this information when selecting benchmarks and performance standards to include in our analysis. Our analysis shows that the data for calendar year 2002 (CY02), the basis for this report, is consistent and reliable.

Selecting Diseases for Analysis

Establishing a disease-specific performance measurement program requires a state to identify a set of diseases for analysis. There are a number of factors to consider when selecting diseases. The individual goals of each state will influence the priority it places on each factor. The following set of decision rules should be considered when making the selection:

- *The incidence of the disease should be reasonably prevalent across the health plans.* A large population size will not only increase the validity of the analysis, but also increase the likelihood that the health plans will take an interest in the results. A state might also choose to consider the frequency of the disease among subpopulations of interest (e.g., SSI enrollees).
- *The disease must have clinical standards of care, and those standards should include process measures that are strong predictors of positive outcomes.* A successful performance measurement program will evaluate health plan performance against a well-defined and accepted standard. We will discuss performance measures in more detail later in the report.
- *The performance of the health plans in treating the diseases must be measurable with available data.* Because this study relies solely on administrative data, each disease selected must have performance indicators that are not dependent on medical chart reviews. It is important, as well, for each state to understand the quality of its data and how data quality might impact the results of the analysis.
- *The potential for quality improvement and cost savings within the Medicaid program should exist.* The feedback provided through a performance measurement program should encourage health plans to improve performance, with the additional goal of improving health care outcomes and providing cost savings, where appropriate.

Our goal with this project was to focus on measuring the performance of health plans in treating chronic diseases. While most Medicaid managed care programs primarily enroll women and children (i.e., the TANF population) and therefore focus their quality

improvement efforts on access measures for these populations, Maryland included its SSI population in managed care from the program's inception.[2] Although SSI enrollees represent only 20 percent of the enrolled population, their costs are 40 percent of annual Medicaid expenditures. Chronic diseases are more prevalent in the SSI population than in the TANF population. Furthermore, we expected chronic diseases to be more sensitive to diagnosis-based risk adjustment.

Our initial step in applying the decision rules listed above was to evaluate the frequency of each diagnosis within the Medicaid population. We wanted to select diseases that affected a relatively large proportion of the health plans' enrollees. We also considered the distribution of the disease across the following subpopulations: TANF enrollees, SSI enrollees, adults, and children. An analysis of the most frequent diseases in the TANF and SSI populations revealed that three of our final diseases (asthma, schizophrenia, and diabetes) were in the top 20 of frequently occurring diseases for both the TANF and SSI populations in CY01.[3]

Once we identified the most frequently occurring diseases, we applied the remaining decision rules. For each disease under consideration, we examined the availability of clinical standards of care, the quality of the data, and the potential for disease management leading to cost savings. We also needed to develop a definition that assured that the population contained only those enrollees with the diagnosis under study, but did not exclude a large number of potentially appropriate subjects. Several diseases were eliminated from consideration because of data concerns, lack of quality of care indicators, difficulty defining the correct population, or size of cohort. The eliminated health conditions/diseases included: pregnancy, substance abuse, childhood seizures, heart disease, attention deficit hyperactivity disorder, hypertension, sickle cell disease, and depression.[4]

The final list of diseases is:

- Asthma;
- Diabetes;
- HIV/AIDS; and
- Schizophrenia.

---

[2] *SSI enrollees under the age of 65 who were not dually eligible were enrolled.*

[3] *Much of our preliminary analysis in selecting conditions was done with CY01 data, as CY02 data was not complete at the time. For this analysis, we measured the frequency of chronic diseases using chronic expanded diagnostic clusters (EDCs). EDCs, developed by Johns Hopkins University, are broad groups of diagnostic codes (see www.acg.jhsph.edu). They represent disease markers. A person may have more than one EDC. States without this added resource may evaluate the frequency of diagnoses.*

[4] *Our reasons for dropping individual conditions highlights some of the problems states will encounter when developing disease-specific performance measures. For example, sickle cell disease had high costs and great interest among members of the Advisory Group, but the size of the cohort was too small to support cross plan profiles. In contrast, attention deficit hyperactivity disorder appeared to have a large cohort, but a lack of clinical consensus as to diagnosis and treatment made performance measurement too difficult.*

Defining Disease Cohorts

One of the most challenging aspects of developing a disease-specific performance measurement program is defining the cohorts to be studied. There are two major criteria to consider when developing a definition: enrollment and clinical.

Enrollment Criteria

How a state defines enrollment criteria will effect what a state is able to measure. It also implicitly conveys a state's expectations of its health plans. For the decisions about enrollment criteria, the trade off is between the size of the population being assessed and continuity of enrollment within a single health plan. Decisions that favor greater continuity within a plan (i.e., longer periods of enrollment with no or only very short gaps in enrollment) will inevitably drop individuals from consideration. Criteria that define away individuals with short or inconsistent enrollment may send health plans the unintended message that it is not important to effectively serve those enrollees. While the argument that health plans should only be evaluated on those individuals for whom they have served for extended periods of time is reasonable, it is also problematic as Medicaid populations tend to have more inconsistent enrollment than commercial populations.

The following questions should be considered as a state develops its enrollment criteria:

- *How will extending or restricting the enrollment requirement affect the size of the cohort?* States must use caution if extending the enrollment period decreases the size of the cohort so that it is too small for meaningful analysis.
- *Do you want to measure individual health plan performance or overall program performance?* The answer to this question may influence whether you restrict your cohort to enrollees who have been enrolled in a single health plan for the period under study. Health plans will resist being measured on the care they have provided to enrollees who have only been enrolled in their plans for the last few months of the period under study.
- *How much time will you give a health plan to become "accountable" for the care provided to its enrollees?* The enrollment requirement for a performance measurement program suggests that during that period of time, a health plan should have delivered the minimum level of services to all of its enrollees.
- *Will you allow gaps in enrollment during the period under study?*
- *Will you define the population in the same year that performance is being measured, or will you pull the cohort for a previous time period?* Selecting the cohort during the previous time period (e.g., a year before the year being measured) provides additional time for which Medicaid and the individual health plans are responsible for the enrollee's care, but restricts the cohort to enrollees with two consecutive years of eligibility. HEDIS® varies its enrollment criteria depending on the measure under study; a state might choose to vary enrollment criteria depending on the disease and/or performance measures being considered.

To define our enrollment criteria, we conducted a sensitivity analysis of the length of the enrollment period for individuals with asthma. Our first iteration compared the number of enrollees with asthma who were enrolled in the managed care program for a minimum of 184 days and a minimum of 320 days, regardless of health plan. Extending the length of the enrollment period decreased the size of the cohort, but increased the length of time for which Medicaid supported the health care provided to the individual. We then added a requirement for enrollment in a single health plan to improve the state's ability to hold a single health plan accountable for the care provided to the individual. Table 1 indicates the number of enrollees with asthma who were continuously enrolled in a single health plan for the minimum number of days in CY02.

Table 1. Number of Enrollees Meeting the Asthma Definition Who Were Continuously Enrolled in a Single Health Plan for Number of Days Indicated (CY02)

Health Plan	184+ Days	320+ Days
MCO A	9,273	7,505
MCO B	911	688
MCO C	9,840	8,050
MCO D	12,259	10,157
MCO E	1,676	1,375
MCO F	10,806	8,645
All MCOs	44,765	36,420

After verifying that it met our goals, we adopted our final enrollment criteria from HEDIS®. Every person in the study was enrolled in the same health plan for at least 320 days and enrolled as of December 31st of the study year. None of the subjects had more than one gap in enrollment during the year, and any gap could not exceed 45 days.

Clinical Criteria

Developing a clinical definition for each disease cohort presents additional challenges. First, a state needs to decide whether to develop broad or narrow definitions. A broad clinical definition would increase the size of the cohort and increase the likelihood that most of the enrollees with the given disease would be included in the study. However, it also increases the likelihood that enrollees who do not have the disease under study would be erroneously included in the cohort because of coding errors or anomalies.[5]

[5] *A commonly cited problem with administrative data is that they do not distinguish a "rule out" diagnosis from an actual diagnosis. For example, an individual presents at a physician's office with a set of symptoms that may indicate a disease. The providers orders tests to confirm or rule out the disease and the disease is recorded on a claim or encounter form as the diagnosis. The subsequent test demonstrates that the possible diagnosis was not correct and it never appears on later claims. Decision rules*

With broad definitions, states risk holding a health plan responsible for providing certain types of care to individuals in the cohort for a disease that they do not really have. Obviously, this may affect the final results. An alternative is to select more narrow definitions. This results in a smaller cohort, but minimizes the number of enrollees in the cohort who do not qualify with the given disease. When monitoring performance on individual diseases, specificity in defining the cohort is more important than size.

There are several sources available for model definitions, including HEDIS®, EDCs (expanded diagnostic clusters), and approaches used in other published studies. The individual characteristics of a state's program might also offer opportunities for definitions. For example, in Maryland, enrollees who have been documented as having an HIV or AIDS diagnosis are placed into separate capitation rate cells. This provided a readily available and reliable cohort definition for our analysis.

Our project team decided to develop a narrow definition for each cohort. We conducted sensitivity analyses for a variety of clinical definitions. Table 2 compares two of the definitions we considered for enrollees with asthma: the HEDIS® definition and a broader definition that included enrollees who have either one asthma diagnosis or one asthma prescription. While the broader definition increased the size of the cohort to 49,689, we decided to apply stricter criteria to minimize the risk of including enrollees in the cohort who did not actually have asthma.

Table 2. Number Of Enrollees Who Meet The Asthma Definition Indicated (CY02)

Health Plan	Any Asthma Diagnosis or Prescription	HEDIS Diagnostic Criteria
MCO A	10,267	5,122
MCO B	1,057	519
MCO C	10,988	4,904
MCO D	13,531	6,129
MCO E	1,955	971
MCO F	11,891	5,477
All MCOs	49,689	23,122

Evaluation of a definition may require several steps. For example, our initial definition of schizophrenia included a large number of children. Recognizing that schizophrenia is not frequently diagnosed in children, we analyzed the frequency of diagnosis codes that were included in the definition being tested. We discovered a disproportionate use of a single diagnosis code that can be applied to diseases other than schizophrenia. With clinical

that accept any diagnosis, anywhere, anytime, are vulnerable to the criticism that they will capture individuals who do not have the condition.

guidance, we eliminated that code from our final definition and reduced the number of children in the final cohort.

As mentioned earlier in the report, the ability to define a cohort for study may affect a state's ability to include a disease in its program. For example, we evaluated several definitions of depression before deciding to eliminate it from consideration. More information about the challenges we faced in defining a cohort of enrollees with depression is included in Section V: Areas for Further Research.

A clinician with expertise in the given clinical area reviewed the final definition for each of the diseases presented in this report. Table 3 indicates the source of the definitions for each of our final study diseases.

Table 3. Source of Definition for Each Disease Condition

Condition	Source
Asthma	HEDIS
Diabetes	HEDIS
HIV/AIDS	Capitation Payment Rate Cell
Schizophrenia	EDCs (modified)

Performance Measures

Once a state has decided which diseases to include in the study and defined the cohort of individuals with each disease, the next step is to select performance measures on which to evaluate the health plans or the entire managed care program. Performance measures:

- *Should be measurable with available data.* The methodology we propose in this report suggests that states use administrative data, which restricts the indicators that can be included. For many chronic diseases, test results are preferred standards for measuring health outcomes and, thus, health plan performance. However, most states will not have access to administrative data that include test results. Collecting medical record data is often an expensive and lengthy process, whereas the use of administrative data is less costly and much easier. In lieu of test results, the fact that the test was administered can be used as an alternative measure. Whatever the data source, it is important to ensure that the performance indicator can be measured with the available information.
- *Should be based on nationally recognized benchmarks, standards of care, or research.* By applying this second criterion, a state will increase the likelihood of acceptance by the health plans and their providers.

A first step in identifying potential measures is to review the literature for each disease. Suggested sources include the National Committee for Quality Assurance (NCQA) HEDIS® measures, clinical standards available through the National Guideline Clearinghouse, other published research standards, and measures adopted by other state agencies. If appropriate data are available, a state should include a combination of process and outcome measures in its analysis. Process measures include procedures or tests that are part of the standard of care for the disease (e.g., a blood test). Outcome measures are results or outcomes after care is provided (or not provided). Common examples of outcome measures include mortality and inpatient admission rates.

States should also consider including a combination of "generic" and disease-specific measures. "Generic" measures include services that are accessed by all enrollees being studied, regardless of the disease. Examples include inpatient admissions and ER visits. Including generic measures in the analysis allows states to compare the results for the same measure across diseases to identify patterns. While states may not have a benchmark (or standard of care) to compare the inpatient and ER visit results to, states can compare health plans to each other.

Including disease-specific measures will allow a state to better assess the care being provided for the disease under study. An example of a disease-specific measure is the hemoglobin test (HbA1c) for diabetes.

The next step is to test the data. States should create test reports for each of the indicators under consideration to identify data problems. For example, if lab data are missing from one of the largest providers, the results for a measure using lab data might be incomplete and biased for one or more of the health plans. Where data anomalies exist, an analysis of two or more years of data may help a state identify the source of the problem and whether it is unique or ongoing.

Once states have developed a list of possible measures to include and tested the data to ensure the validity of the results, they should then go through a process of evaluating what to expect from the results. The following questions should be considered when developing the final list of measures:

- Is the measure part of the standard of care for this disease?
- Would we expect health plan performance to vary based on the health status of the enrollee (i.e., is it sensitive to health-based risk adjustment?)?
- Should we include measures that are indicated in the standards of care for the disease only under certain clinical conditions (e.g., p.r.n.[6])?
- How should we interpret the results for "p.r.n. measures"?

As indicated earlier, measures should be based on nationally recognized benchmarks or other sources of clinical standards of care. An ideal indicator might be one that every

[6] *P.r.n. or pro re nate is a commonly used clinical term for "as needed."*

person in the cohort should have (e.g., HbA1c test for diabetes). There is little debate among the clinical community about this indicator. People with diabetes should have the test regardless of how severe the disease (or other co-morbid conditions) may be. For most indicators, states will be able to identify the clinical standard (e.g., administering a test two times per year); the performance standard being measured (which might be lower than the clinical standard depending on the enrollment criteria and other parameters of the performance measurement program); and the percentage of enrollees who meet the standard. A state may choose to evaluate each plan against the performance standard, or against other health plans.

The use of "p.r.n. measures" or measures for tests that are only part of the standard of care under certain clinical conditions introduces an additional level of complexity to the analysis. If a state is unable to measure the clinical conditions under which the test should be administered, the results must be evaluated with caution. Interpreting the results will be easier if administration of the test is expected to increase with the severity of health status, and the health-based risk adjustment tool being used accurately measures that level of severity. For most p.r.n. measures, however, the clinical standard and the performance standard are not clearly defined. In these cases, states can evaluate the performance of each health plan against the others to identify plans whose performance results differ from the norm. States must then consider whether the outliers represent care that is better than the norm, or worse. The use of p.r.n. measures is not impossible, but requires considerable additional analysis to ensure that the results are being interpreted correctly.

We began our process of identifying possible performance indicators by reviewing the literature for each disease. The next step was to test the data for each of the indicators to determine whether we had the data necessary to measure the standard. Although we expected the data to be available for certain indicators, we discovered that the quality of the data was not always as good as expected.[7] Hence, we modified the definitions. In some cases, we needed to expand our definition to include local codes. In other cases, we modified the HEDIS® standards to adjust for the fact that we did not have access to medical records data.[8] We evaluated each measure individually to decide if we should include it in our analysis given the modifications that would be necessary. In some cases, we had to eliminate measures from consideration due to lack of reliable data.

For each disease, we identified a set of disease-specific measures on which we could evaluate health plan performance. We also developed the following set of generic measures that were applied to all four disease cohorts. We measured the percentage of enrollees in the cohort during CY02 who had at least:

---

[7] *Two of the standard lab tests performed on enrollees with HIV/AIDS are CD4 test and viral load. The data from health plans that rely on large hospital-based provider groups for HIV care showed much lower testing rates than other plans. We discovered that the lab data on outpatient hospital claims do not include the type of lab test performed.*

[8] *For example, the HEDIS measure for eye exams for enrollees with diabetes allows health plans to include exams performed in the prior year if the enrollee meets certain criteria. The criteria include lab values and a negative diagnosis of retinopathy indicated in the medical record. Because we did not have access to these data, we were unable to include eye exams performed under those conditions in our analysis.*

- One inpatient admission;
- One ER visit;[9, 10] and
- Two ambulatory care visits.[11]

We have a high level of confidence in the data for these measures and selecting standard measures allows us to compare the results for the same measure across disease conditions to identify patterns. While we do not identify a benchmark (or standard of care) to compare the inpatient and ER visit results to, we do compare the performance of each health plan to the others. We also risk adjust the results where appropriate.

We reviewed our final list of performance measures with a clinician who had expertise in the disease. Careful review from a local clinician serves several purposes. In addition to verifying the accuracy of the definitions, the clinician will call attention to coding concerns (e.g. commonly used rule-out codes) that may affect results. Furthermore, a local clinician will be familiar with local practice patterns that may impact a state's results.

Empirical Evidence for Selecting Performance Measures

Selecting which indicators to include in a performance measurement program is an important decision for a state. Health plans tend to focus resources on the areas on which they will be measured in order to improve their scores. States want to select measures that will meet several goals, including improving care to the population and reducing unnecessary expenditures. An additional feature of this report is an analysis of the demographic and utilization factors that are associated with improved medical outcomes. The results of this analysis will help states select measures that support their goals. While administrative data are a poor source of information for measuring medical outcomes, there are utilization measures, such as inpatient admissions, that can serve as useful indicators of medical outcomes. Most state policy makers and health plans would agree that reducing inpatient admission rates is a positive outcome. In turn, states should select performance measures that support that desired outcome.

In an effort to assist states with identifying the factors could be associated with a desired outcome, we performed a logistic regression for each disease. We considered the impact of demographic characteristics and the attainment of certain disease-specific process measures on the likelihood of an enrollee having an inpatient admission. While demographic characteristics are out of a health plan's control, plans can influence their performance on process measures. We found that having two or more ambulatory care visits during the year decreased the likelihood of an inpatient admission by a third for

---

[9] *An ER visit is defined as a visit to an emergency room that does not result in an inpatient admission.*

[10] *The authors recognize that ER utilization is not necessarily a negative outcome. It is included in this report to demonstrate the application of risk adjustment to a performance measure that can be easily identified using administrative data. States may choose to conduct additional research on ER utilization patterns in their managed care programs before adopting it as a performance measure.*

[11] *An ambulatory care visit is defined as a visit to an outpatient hospital department, a health clinic, or a physician's office.*

enrollees in the asthma, diabetes, and HIV/AIDS cohorts. For schizophrenia, enrollees who had two or more mental health visits during the year were less likely to have an inpatient admission for a medical condition than enrollees who had less than two visits. Our empirical evidence supports the conceptual argument that ambulatory care visits (which are less costly) represent "good" care and are associated with a decreased likelihood of more costly inpatient admissions. A more complete discussion of our methodology and results can be found in Appendix II.

The Advisory Committee

When designing a performance measurement program that will evaluate the performance of specific health plans, it is important to gain acceptance from the health plans that will be evaluated. States should consider including health plans in the planning process to find out which diseases they have been focusing on, where they would like to improve performance, and whether they have concerns about the appropriateness of the measures being considered.

To that end, we formed a Project Advisory Committee that provided input at several points throughout the duration of the project. Members of the committee represented the health plans participating in the Maryland Medicaid program, the Maryland Health Care Commission, the Maryland Special Needs Advisory Committee, Maryland hospitals, and consumer groups. Committee members included physicians, directors of quality improvement activities, and chief executive officers.

The committee met four times during the course of the project to provide critical feedback on the decision rules regarding disease identification, cohort definitions, performance measures, and risk adjustment methodology. Based on their suggestion, we included HIV/AIDS in our analysis. They also helped refine the list of diseases under consideration. For example, we eliminated attention deficit hyperactivity disorder after committee members suggested that a lack of clinical consensus as to diagnosis and treatment would make performance measurement difficult. Committee members offered valuable feedback on whether the performance indicators under consideration were consistent with practice patterns in Maryland. The health plan representatives shared findings from their own quality improvement analyses. Appendix I includes a list of the committee members and the meeting dates.

Interest to Other States

As states develop systems designed to reward or penalize health plans based on performance, they must confront a series of related choices: what should we measure; how should we calculate measures; what is the quality of the data available; how should the performance targets be established; what are reasonable performance expectations; and what financial rewards or penalties should be assessed based on performance for various measures? Faced with these decisions, state performance measurement strategies have tended to use indicators that are widely recognized and enjoy a broad consensus of

support. Hence, HEDIS® or HEDIS®-like measures have tended to form the backbone of state performance programs.

As was discussed earlier, HEDIS® measures, while widely recognized and well accepted, do not extensively address issues related to the care of chronically ill populations. This limitation is not especially significant when state programs are focused largely on populations that are not chronically ill, such as TANF eligibility categories. However, states are increasingly interested in applying managed care principles to serve individuals with disabilities, who tend to have more chronic diseases. When comparing health plan performance on the care of chronic conditions, it is particularly important to focus on outcome measures whenever possible. States trying to develop their capabilities in this area can use this report to advance their ability to effectively develop and incorporate performance measurement into their evaluation and payment systems in several ways.

- **Identifying and selecting process measures.** Implicit in most state approaches to performance measurement is the belief that the provision of "good" care (i.e., the provision of certain services at the appropriate time) will lead to improved outcomes for the population served (e.g., better health status and reduced hospitalizations). A challenge for states is how to choose between an array of process measures available. The possible measures vary in both their breadth (the portion of the population they apply to) and their complexity to calculate. The analysis offers states a useful criterion for selecting among process measures: a measure's contribution to lowering hospital admission rates. There are measures, however, that states may want to select that do not necessarily lead to lower hospital admission rates. For example, access to dental services may be an important element in a state's performance measurement program but will not significantly impact inpatient admissions rates. Still, the criterion of lowering hospital admissions does offer states an additional strong rationale for choosing among and justifying competing measures.

- **Performance measurement needing risk adjustment.** The analysis also addresses an important issue that states may face when assessing health plan performance: determining what measures are appropriate candidates for risk adjustment. Some states that have sought to impose penalties upon health plans based on poor performance on specific process measures have faced criticism for not risk adjusting those measures. Health plans have argued that since population risk profiles vary, performance indicators, including process measures reflecting agreed standard of care, should adjust for that variation. The results presented in this report indicate that for the process measures assessed (e.g., ambulatory visits and lab tests), that argument does not hold.

  On the other hand, outcome measures such as inpatient admissions and ER use are sensitive to variation in the risk profiles of health plans and are good candidates for risk adjustment. In theory, both inpatient admissions and ER use should decline with managed care, as these systems should emphasize outpatient

services and preventive care. It is reasonable, therefore, for states to include such indicators in their performance measurement programs. Because both of these services are sensitive to risk, the establishment of a standard benchmark for inpatient admissions or ER use is an untenable task. States who lack either the data or the resources to perform risk adjustment, therefore, should not consider using either inpatient admissions or ER use as part of a performance measurement program.

States that do possess the resources to conduct risk adjustment have several options for applying it to performance strategies. Risk adjustment can be used, as it is in this analysis, as a way to assess a health plan's performance by comparing it to the performance of other participating plans. A health plan that exceeds the risk-adjusted standard for a given condition could be assessed a financial penalty. Conversely, a health plan that performed well could be rewarded (assuming it scored well on relevant process measures).

States with well established risk adjustment systems might wish to consider going a step further, by building an efficiency adjustment into the development of payment rates. An efficiency adjustment would use risk adjustment as a tool for identifying and removing costs that reflect poor management on the part of some plans from overall payment rates. Often in the calculation of payment rates, the combined health plan reported costs are used as a starting point. Thus, inefficient plans tend to drive up the overall average cost. The analysis suggests that risk adjustment techniques could be used by states to refine their payment mechanisms to more precisely reward efficient plans, penalize inefficient plans, and capture savings for the state.

Policy Implications

Developing a disease-specific performance measurement program requires a significant investment of time and resources for any state. Risk-adjusting the results add a layer of complexity and investment. While the endeavor may yield valuable results, states should also consider the policy implications of pursuing this path.

Any focus on disease-specific performance measures requires states to understand the data they are using and how data anomalies may affect the results of the analysis. Administrative data are generally perceived to be a poor source of information for measuring medical outcomes. At the same time, administrative data can be a valuable source for information about process measures. Payment mechanisms (such as capitation payments paid to clinicians) and claims coding issues (such as the use of a single code for all lab tests) may impact the type of measures a state can effectively produce. Also, when selecting which diseases to analyze, it is important to remember that interest in the disease may change over time. As with any measurement program, states run the risk that heath plans will focus too many resources on the disease that is being analyzed, to

the detriment of other disease conditions. Selecting diseases that are prevalent and offer the opportunity for quality improvement and cost savings is important.

The application of risk adjustment can lead to more accurate measurement of health plan performance and yield information that helps health plans address deficiencies (e.g., a health plan whose inpatient admission rate remains high even after adjusting for risk). Applying risk adjustment, however, also has policy implications.

Once a state introduces risk adjustment to performance measurement, health plans may expect or demand it for certain measures. States should carefully consider how they approach the decision rules impacting risk adjustment. Health plans that perform poorly on certain benchmarks may argue that, if adjusted for risk, their scores would be better. Investing the resources to risk adjust every measure is not something that most states should or want to do. Existing performance measurement efforts in most states focus on preventive services such as well child visits and vaccinations. These are clearly not appropriate for health-based risk adjustment. In contrast, our results suggest that performance measurement can improve the accuracy of some outcome measures. As states expand the focus of their measures, developing guidelines for evaluating the appropriateness of risk adjustment will be important.

States that decide to pursue risk adjustment must also decide what factors to include in an adjustment. Demographic factors such as age, sex, region, and eligibility category are most frequently included. The paper introduces the application of an additional variable – health status. The availability of the data necessary will often determine what factors a state can include in an adjustment.

Finally, while risk adjustment can improve performance measurement, it is more difficult to understand. Many state Medicaid programs produce consumer report cards that include performance measures for the participating health plans. States will want to carefully evaluate which measures and techniques are used to develop consumer materials and which are used exclusively for internal reporting and quality improvement feedback to plans.

Section II: The Role of Risk Adjustment

When comparing health plan performance, differences in case mix across health plans can impact the final results. Risk adjustment, a method of statistically adjusting for pre-existing characteristics of enrollees, enables us to control for the variation in case mix across health plans, thus strengthening the validity of our results by enabling a more accurate comparison.

What is Risk Adjustment?

Risk adjustment is a process by which the health status and other demographic characteristics of an enrolled population are taken into consideration when assessing

provider performance or determining payments. Risk adjustment helps users understand and appreciate differences in physician resource use across practices or plans. It is also an important consideration in establishing equitable premiums, capitation rates, withhold returns, or other forms of budgeted payments. Risk adjustment, in the application to management or financing of health care services or profiling, incorporates the health characteristics of patients, as well as their propensity to use clinical services. Because these factors usually cannot be measured directly, characteristics such as age, gender, and diagnosis are commonly used as proxies for health status.

An effective disease-specific performance measurement program can be developed without the use of risk adjustment, using the guidelines proposed in Section I. However, states that have the capacity should consider incorporating risk adjustment into their performance measurement programs. Risk adjustment can provide a fairer assessment of differences in expected performance across health plans. For example, there are measures, such as inpatient admissions, for which performance can be expected to vary based on the health status of the population. That is, a health plan with a sicker case mix might be expected to admit more enrollees than a health plan with a healthier case mix. Risk adjustment provides a method of controlling for that expected difference, allowing states to better assess a health plan's performance given the case mix of its population. This enables states to better evaluate health plans on the basis of quality and efficiency, while reducing the impact of factors beyond the plans' control.

The application of health-based risk adjustment to performance measurement is relatively new in Medicaid. In the past, many states have used risk adjustment in order to take diagnosis-based health status into account when setting and allocating capitation budgets for providers in Medicaid managed care programs. This has been an advance from earlier adjustment methods based on demographics, eligibility type, and geographic factors, or simply prior spending. Much of the enthusiasm for this application of risk adjustment to payment policies is driven by the notion that if payments can reflect (and thus better match) the actual resource needs (i.e., the risk assumed by managed care plans) of a given population, the incentives for providers to avoid high-cost persons or groups would be mitigated, thereby reducing gaming of the payment system and encouraging a more equitable and fair system.

Provider profiling and the monitoring of health plan performance is a natural extension of applied risk adjustment because health status, and thus risk, is not randomly distributed across plans or providers. Making adjustments for varying health status for more accurate and apt comparisons of provider performance is just as important as it is for determining provider payment.

How Does Risk Adjustment Work?

When is it appropriate to apply risk adjustment to performance measurement? In general, the measure of interest should vary significantly across the risk factor of interest, and the risk factors should vary across the entities subject to profiling. Moreover, data

should be readily available and well populated among the providers. In short, the data should be robust enough so that the individual impact of a health plan can be determined. A more detailed discussion of the factors that influence whether to risk adjust a performance measure is presented at the end of this section.

Since the inception of the Medicaid managed care program, Maryland has applied diagnosis-based risk adjustment using the Adjusted Clinical Group (ACG) case mix system to determine capitation payments to health plans. The analysis presented here applies the same risk adjustment methodology to performance measurement in Maryland's program.

The ACG case mix system assigns each enrollee to a single ACG category according to his individual pattern of morbidity based on diagnostic information available through claims and encounter data. Although Maryland carves out its mental health services from the managed care program, the diagnostic information from mental health claims is used in assigning enrollees to ACGs. To avoid working with 82 independent ACGs, we grouped them into six categories, or Resource Utilization Bands (RUBs), according to similar expected cost and morbidity. The six RUBs, presented in increasing levels of morbidity, are: Non-Users,[12] Healthy Users,[13] Low Morbidity, Moderate Morbidity, High Morbidity, and Very High Morbidity.

The following example illustrates how risk adjustment can be useful when assessing the performance of health plans in providing care to enrollees with asthma. As mentioned earlier, the number of enrollees who are admitted to the hospital could be expected to vary across health plans according to the case mix of each plan's enrolled population. This example considers the percentage of enrollees with asthma who have had at least one inpatient admission during the calendar year.

When deciding whether to risk adjust a performance measure, states should first consider whether the measure varies by the risk factor of interest (morbidity). Table 4 presents the percentage of enrollees with asthma who had at least one inpatient admission, according to their RUB (our measure of morbidity). The data show that the percentage of enrollees with at least one admission increases with RUB severity, from a low of 3.1 percent in the Low Morbidity group to a high of 64 percent in the Very High Morbidity group.[14]

---

[12] The Non-Users RUB includes members of the cohort who do not have enough diagnostic information on their claims/encounter data to be accurately classified into the appropriate risk strata. For example, an enrollee may qualify as a member of the asthma cohort by filling an asthma prescription at some point during the year. However, prescription information is not used by the ACG system to assign enrollees to ACGs/RUBs. Therefore, if an enrollee only received prescriptions, and does not have any diagnosis information during the year, he would be a member of the Non-Users RUB.

[13] The Healthy Users RUB includes enrollees whose diagnostic information contains only data about preventive services and minor conditions. The data are not sufficient to accurately classify the enrollee into the appropriate risk group.

[14] The results for the first two morbidity groups (Non-Users and Healthy Users) are zero because documentation of an inpatient admission in the administrative data would place any enrollee with asthma into at least the Low Morbidity RUB.

Table 4. Percent of the Asthma Cohort With at Least One Inpatient Admission by RUB (CY02)

RUB	Percent of Cohort
Non-Users	0.0
Healthy Users	0.0
Low Morbidity	3.1
Moderate Morbidity	7.7
High Morbidity	34.0
Very High Morbidity	64.0
All RUBs	17.7

The next step is to consider whether the risk factor of interest (morbidity) varies by the unit of analysis (health plan). That is, what is the case mix distribution of enrollees across health plans? Do some plans have a greater percentage of enrollees in the Very High and High Morbidity RUBs than other plans? If the percentage of enrollees with an admission increases with morbidity, as we demonstrated in Table 4, then a health plan with a higher morbidity index is likely to have a greater percentage of enrollees with an admission than a health plan with a healthier case mix. This difference in admissions might be appropriate given the case mix of each plan.

Table 5 presents the case mix distribution for the asthma cohort for each health plan. The percentage of asthma enrollees in the Very High Morbidity RUB ranges from a low of 8.2 percent (MCO F) to a high of 18.8 (MCO B). The distribution in the High Morbidity RUB ranges from a low of 18.3 percent (MCO C) to a high of 32.8 percent (MCO B). Similar patterns are evident in the Low and Moderate Morbidity RUBs, illustrating that the case mix does vary across the entities being profiled (health plans).

Table 5. Distribution of Asthma Cohort Across Health Plans by RUB (CY02)

RUB	MCO A	MCO B	MCO C	MCO D	MCO E	MCO F	All MCOs
Non-Users	1.3	0.6	2.0	2.0	0.9	2.0	1.7
Healthy Users	1.3	0.6	2.1	1.6	1.2	1.9	1.7
Low Morbidity	15.3	6.7	19.0	20.0	15.5	21.0	18.6
Moderate Morbidity	48.6	40.5	49.7	45.4	46.4	46.7	47.3
High Morbidity	21.8	32.8	18.3	20.2	21.0	20.3	20.5
Very High Morbidity	11.6	18.8	8.9	10.7	14.9	8.2	10.2
All RUBs	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 6 presents the inpatient admission data stratified by RUB for each health plan. Stratifying the data by RUB allows us to compare enrollees with similar morbidity across health plans. As was the case statewide, the percentage of enrollees with an admission increases with the severity of the RUB within each health plan. A table like this makes it easier for states to compare health plan performance for similar types of enrollees, but comparing six numbers for each performance benchmark is cumbersome.

Table 6. Percent of the Asthma Cohort With at Least One Inpatient Admission (CY02)

RUB	MCO A	MCO B	MCO C	MCO D	MCO E	MCO F	All MCOs
Non-Users	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Healthy Users	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Low Morbidity	2.3	8.7	1.6	4.9	0.0	3.0	3.1
Moderate Morbidity	8.9	7.3	6.4	9.0	10.5	6.1	7.7
High Morbidity	37.3	17.9	30.2	35.6	30.4	34.6	34.0
Very High Morbidity	64.2	46.9	60.6	67.0	68.4	64.0	64.0
All RUBs	20.3	18.2	14.4	19.4	21.5	15.7	17.7

Applying logistic regression techniques provides a method of calculating an expected admission rate, resulting in one number per health plan, rather than six. With logistic regression, one can control for variables other than health status that might affect performance results, such as demographic characteristics. The power of multivariate regression analysis is that all of these factors can be controlled simultaneously.

For what additional factors might a state choose to control? States should consider only those factors that are independent of health plan control but that might affect performance results. Demographic factors are most frequently included. The analysis presented here includes four characteristics: eligibility category, age, sex, and region. Independent analysis has demonstrated that service utilization varies across each of these factors and conceptual arguments support this theory.

A logistic regression can be used to demonstrate the relationship between a dependent variable (the measure of interest) and any independent variables (risk factors) that may influence performance on the measure. In the following analysis, a regression is used to relate enrollees' inpatient admissions to their RUB (as a measure of health status) and the four demographic characteristics mentioned above. Dummy variables were created for each variable. The Non-Users and Healthy Users RUBs were combined with the Low Morbidity RUB, resulting in variables for only four RUBs.[15]

The results of the regression analysis are then used to predict the likelihood that an enrollee would be admitted to a hospital. These probabilities are averaged over all of the

[15] *The sizes of the Non-Users and Healthy Users RUBs were too small for meaningful analysis.*

enrollees in a health plan to predict the likelihood that the average enrollee would be admitted to the hospital. These predicted values are then compared to the observed rates for each health plan. This comparison is used to determine whether a plan is performing better or worse than expected given the risk and demographic profiles of its enrollees.

Table 7 contains each health plan's performance results on the inpatient admission measure, both before and after risk adjustment. The first column of data reports the unadjusted percentage of enrollees in each health plan that had at least one inpatient admission. The second column divides the observed value (from the first column) by the statewide admission rate (17.7). Health plans with a score greater than 1.0 have a greater percentage of enrollees with an admission than the statewide average, while those with a score less than 1.0 have a lower percentage of enrollees with an admission.

Table 7. Observed vs. Expected Performance by Health Plan Measure: Inpatient Admissions for Asthma Cohort (CY02)

	Unadjusted Results		Risk-Adjusted Results	
Health Plan	Observed	Observed/ State Avg	Case Mix Expected	Observed/ Expected
MCO A	20.3	1.15	19.4	1.05
MCO B	18.2	1.03	30.0	0.61
MCO C	14.4	0.81	15.4	0.94
MCO D	19.4	1.10	17.7	1.10
MCO E	21.5	1.21	20.3	1.06
MCO F	15.7	0.89	16.8	0.93
All Plans	17.7	1.00	17.7	1.00

The third column shows the percentage of enrollees with an admission that one would expect each health plan to have, given its risk and demographic profile, as predicted by the regression model. The final column compares the unadjusted results (column one) with the expected results (column three). How does the health plan perform compared to how we would expect it to perform, given its demographic and risk characteristics? Again, results greater than 1.0 indicate that the percentage of enrollees with an admission exceeds the statewide average, while scores less than 1.0 indicate that the percentage of enrollees with an admission is less than the state average.

In this case, risk adjustment did influence the performance results for several of the health plans. For example, the unadjusted analysis shows that MCO B had 3 percent more enrollees with an admission than the statewide average (1.03 vs. 1.00). Once adjusted, MCO B's results were much better than the state average (0.61 vs. 1.00). The percentage by which MCO E exceeded the state average decreased (from 21 percent to 6 percent) with risk adjustment. Because there is no change in the results for MCO D, it

appears that the percentage of enrollees who were admitted to the hospital was equal to what would be expected given MCO D's case mix. The results in the final column more accurately reflect the performance of each health plan, taking into consideration the age, sex, region, eligibility category, and health status of its enrolled population.

As mentioned earlier, not all performance measures are appropriate for risk adjustment. In general, states would not need to risk adjust process measures that are part of the standard of care for a disease. If performance on a measure is not expected to vary based on the risk factor of interest, there is little argument for risk adjusting. For example, NIH guidelines suggest that enrollees with diabetes receive a hemoglobin A1c test at least twice a year, regardless of the severity of the person's condition. Risk adjusting for health status should have little influence on the performance results. The data supports this assertion. Tables 8 and 9 include the performance results for hemoglobin A1c testing for a cohort of enrollees with diabetes. Table 8 suggests that the percentage of enrollees who received a hemoglobin A1c test does not vary according to the severity of health status. These results are reinforced in Table 9. There is little change in the performance results for most of the health plans after risk adjustment.

Table 8. Percent of the Diabetes Cohort with at Least One Hemoglobin Test (HbA1c) (CY02)

RUB	MCO A	MCO B	MCO C	MCO D	MCO E	MCO F	All MCOs
Non-Users	15.0	0.0	9.7	4.4	0.0	3.6	7.4
Healthy Users	16.7	-	0.0	0.0	0.0	50.0	14.3
Low Morbidity	60.0	40.0	55.6	40.3	57.1	55.2	51.4
Moderate Morbidity	67.2	67.8	70.0	51.7	72.4	71.2	64.7
High Morbidity	61.6	56.3	68.8	51.2	68.9	71.2	62.3
Very High Morbidity	56.6	71.9	69.3	51.5	68.3	72.1	61.7
All RUBs	61.2	64.7	67.7	50.5	68.0	69.6	61.6

Table 9. Observed vs. Expected Performance by Health Plan Measure: Hemoglobin A1c Test for Diabetes Cohort (CY02)

	Unadjusted Results		Risk-Adjusted Results	
Health Plan	Observed	Observed/ State Avg	Case Mix Expected	Observed/ Expected
MCO A	61.2	0.99	59.9	1.02
MCO B	64.7	1.05	60.7	1.07
MCO C	67.7	1.10	64.0	1.06
MCO D	50.5	0.82	60.6	0.83
MCO E	68.0	1.10	62.0	1.10
MCO F	69.6	1.13	62.8	1.11
All Plans	61.6	1.00	61.6	1.00

Decision Rules Affecting Risk Adjustment

States that are considering risk adjustment in their performance measurement programs face two critical decisions:

- Which measures are appropriate for risk adjustment?
- For which risk factors should the state control?

As demonstrated earlier, risk adjustment is not appropriate for all performance measures. When applied appropriately, risk adjustment can improve a state's ability to assess health plan performance by taking into consideration the case mix of its population. What guidelines, then, should a state follow when deciding whether to risk adjust performance results? Our analysis suggests that it is useful to risk adjust when:

- *The measure varies significantly across the risk factor of interest.* Risk adjustment will have little or no effect on performance results unless there is some variation in the results across the risk factor of interest. For example, if the rate of hospital admissions is the same for all enrollees regardless of their relative health status, then risk adjustment is not useful.
- *The risk factors vary by unit of analysis (in this case, health plans).* If the case mix of each health plan is the same, then there is no reason to adjust the results.
- *There is a conceptual argument supporting the need to identify the "independent" impact of the health plan on care.* If the standards of care suggest that all individuals with a disease should have a certain test, it would be difficult for a state or a health plan to argue that the performance results should be adjusted for health status, regardless of the level of variation.

In most situations, outcome measures such as mortality and hospitalization are appropriate for risk adjustment. In contrast, process measures that reflect a standard of care are not appropriate. As demonstrated earlier, and as expected, hospitalization would occur more frequently among groups who have higher morbidity, yet people with diabetes who have a higher morbidity are not the only ones who should receive an annual hemoglobin (HbA1c) test.

States that decide to pursue risk-adjusted performance assessment must also decide which factors are appropriate to include in their risk adjustment methodology. Demographic factors such as age, sex, region, and eligibility category are most frequently included. States may also include health status. By including a variable in a risk adjustment, a state is implicitly conveying that any variation in performance across health plans because of that factor is acceptable and that the plans are not expected to overcome that variation through outreach, case management, or other means. In general, states should adjust for only those risk factors that are independent of health plan control and that might have a direct impact on plan performance.

Section III. Overview of Results

We applied the decision rules and statistical techniques described earlier to evaluate the performance of six health plans in Maryland's Medicaid program in CY02 on their ability to provide appropriate care to enrollees with four diseases: asthma, diabetes, HIV/AIDS, and schizophrenia. We selected a group of disease-specific measures for each disease in order to evaluate each health plan's ability to meet standards of care. Examples of disease-specific measures include the percentage of enrollees with diabetes who had a hemoglobin A1c test, and the percentage of women with HIV/AIDS who had a pap test during the year. We also selected three generic measures to evaluate the care provided across all four diseases, allowing us to look for patterns in service utilization. The generic performance indicators measure the percentage of enrollees who had at least one inpatient admission, at least one ER visit, and at least two ambulatory care visits during the year. Our analysis included both process and outcome measures.

Following the decision rules presented in Section II, we applied risk adjustment to the performance results, where appropriate. Our analysis suggests that health-based risk adjustment is important when assessing performance on the two generic outcome measures: the percentage of enrollees who had at least one inpatient admission and the percentage of enrollees who had at least one ER visit.

For each disease, we also applied regression techniques to identify process measures that, when controlling for other factors, appear to be associated with a decrease in the likelihood of an inpatient admission. This analysis provides a foundation for the selection of indicators that a state may consider including in a performance measurement program.

Comprehensive results for each disease are presented independently in Appendix II, allowing the reader to select specific diseases of interest. A summary of key findings is offered below.

Summary of Key Findings

- *There is a strong direct relationship between health status (as measured by RUB severity) and utilization rates for inpatient admissions and ER visits*. These results suggest that health plans with a sicker case mix would be expected to have a higher percentage of enrollees with inpatient admissions or ER visits. Applying health-based risk adjustment to the results for these indicators improves the accuracy of the measurement by controlling for any variations in case mix across the plans.

- *Performance on process measures is generally not sensitive to health status*. These results are consistent with clinical expectations. Standards of care that are appropriate for all enrollees with a certain diagnosis should be applied consistently regardless of health status. Applying health-based risk adjustment is not suitable for such indicators.

- *Receiving ambulatory care services is associated with a decrease in the likelihood of an inpatient admission*. When controlling for other factors (including health status), enrollees who had two or more ambulatory care visits were approximately one-third less likely to have an inpatient admission than those enrollees who had fewer than two ambulatory care visits. This conclusion is consistent with the literature on ambulatory care sensitive conditions, which suggests that appropriate outpatient care can reduce the need for inpatient admissions for certain health conditions. Most states and health plans would agree on a goal of reducing inpatient admissions. Hence, this analysis further supports evidence that ambulatory care visits can decrease an enrollee's likelihood of admission, thus supporting the inclusion of an ambulatory care visit threshold in any state's performance measurement program. Even states that do not have sophisticated data systems to stratify enrollees by morbidity or to risk adjust performance results can likely document ambulatory visits.

Summary of Plan Performance

One of the advantages of applying generic measures to several diseases is the opportunity to identify trends in health plan performance across diseases. With such information, states can identify whether some health plans use the ER more frequently than others. States can also identify plans that have consistently low ambulatory care visit rates and high inpatient admission rates, suggesting a need for more focused attention on primary care services.

The following patterns were noted from our analysis of health plan performance in CY02:

- For medical inpatient admissions, only one health plan performs better than average for all four diseases, after applying risk adjustment.
- For ER visits, two health plans perform better than average and two perform worse than average for the three[16] diseases analyzed, after applying risk adjustment.
- On the measure of two or more ambulatory care visits, two of the health plans perform better than average and one performs worse than average for all four diseases.

The results from the disease-specific process measures provide some insight into health plan performance for individual disease. The results show that:

- Three of the health plans perform better than average for diabetes and one health plan consistently performs worse than average.
- One health plan performs better than average on measures for HIV/AIDS and two plans consistently perform worse than average.
- There were no consistent patterns of performance for asthma- and schizophrenia-related indicators.

Detailed information describing the performance of all six health plans on treating enrollees with each of the four diseases is available in Appendix II.

Section IV: Study Strengths and Limitations

One strength of this study is that it reviews high quality data from Maryland's Medicaid managed care program for well-characterized diseases. Maryland's administrative data set has continually improved since the inception of the managed care program in 1997, providing a consistent and reliable source of information about the service utilization for enrollees, thereby enhancing the validity of the results. The strength of the data set is due in large part to the fact that Maryland uses its encounter data to risk adjust the capitation payments to plans, creating a strong incentive for health plans to submit complete data.

Our analysis does have some limitations. First, we used CY02 data for every aspect of the analysis: defining the cohort with the disease, measuring the performance of the health plans in treating the disease, and assigning enrollees to RUBs based on the diagnosis information. Evaluating performance in the same year that one selects the cohort can impact the results. For example, for most of the diseases we analyzed, an enrollee became a member of the cohort by having an inpatient admission, an ER visit, or an ambulatory care visit with the diagnosis of the disease. When the service utilization thresholds in the

[16] *We did not risk adjust ER visit results for schizophrenia.*

definition are also used as performance measures, health plans are "guaranteed" to meet the standard. States should ensure that results are interpreted correctly if the performance measure and cohort definitions are congruent and are applied in the same time period. Our analysis of HIV/AIDS is not subject to this limitation, as the cohort was defined based on payment rates rather than service utilization thresholds.

One must also be cautious about results when risk assignment is done in the same year that performance is measured. For example, an enrollee must seek a certain intensity of service in order to be classified in a more severe RUB. That may, by definition, allow the enrollee to meet the utilization thresholds being measured. One method of reducing the influence of this circularity is to use two periods of data. Data from the earlier period can be used to evaluate health status and assign enrollees to RUBs; data from the latter period can be used to measure performance. This may also reduce the number of enrollees whose data are insufficient to be properly categorized into RUBs other than Non-Users or Healthy Users, which would improve the overall analysis.

An additional concern is that our analysis is not based on an episode of care. It is possible for an enrollee to be diagnosed with the chronic disease under study during the last month or two of the year. This limits the amount of time a health plan would have to meet the performance standards being measured. However, this issue is not of great concern, as it is likely that this "disadvantage" would be evenly distributed across health plans.

Finally, while we were able to draw large enough cohorts for each disease to support meaningful analyses on a statewide level, there are some diseases for which the individual health plan numbers are fairly small. For example, in HIV/AIDS, the distribution of enrollees is disproportionate to one health plan. As a result, the statewide means are heavily influenced by the performance of that one plan. Hence, the results should be interpreted with caution.

Section V: Areas for Further Research

The analysis presented in this report demonstrates how risk adjustment can improve a state's ability to evaluate health plan performance in a Medicaid program. As with any study, there are areas where further research is recommended. First, our analysis uses the performance threshold of two ambulatory care visits per year. While this threshold was based on clinical standards for some of the diseases we studied, this may not be the appropriate threshold for all diseases. Would a different threshold change the relationship between ambulatory care visits and reduced inpatient admissions? Do more ambulatory care visits during a year result in an even greater likelihood that an enrollee will not have an inpatient admission? Does the correlation between ambulatory visits and reduced likelihood of an admission hold for enrollees who only have one visit per year?

Similarly, the threshold of one ER visit as an outcome measure may not be appropriate for all diseases. ER utilization is not universally accepted as a negative outcome, as expressed by some of the health plans representatives on our Advisory Board. Multiple ER visits during the year might be a better outcome measure, suggesting "inappropriate" ER use (e.g., substituting an ER visit for an ambulatory care visit). Additional analysis could improve the selection of an appropriate threshold.

The results for schizophrenia demonstrate that continuity of medication is related to the likelihood of an inpatient admission. Drugs that treat mental health conditions are often exempt from oversight mechanisms (such as preferred drug lists) employed by states to control costs. At the same time, they remain one of the most expensive categories of drugs reimbursed by Medicaid. Additional analysis of prescription drug utilization, including appropriate dosages and refill patterns, might yield some information for health plans and states to use to better address the issues of cost and appropriate utilization.

Several diseases were eliminated from consideration in our analysis for various reasons, including the size of the cohort, the lack of quality of care indicators, and data concerns. Specifically, we eliminated a health condition that receives considerable interest—depression—because defining the cohort proved problematic. As described in Section I, defining the cohort of enrollees with the disease under study requires precise decision rules. We also cited that "specificity" (i.e., did every member of the cohort actually have the disease?) was a primary value. When attempting to define depression, we had difficulty meeting this criterion. With a broad definition, we ran the risk of including enrollees in the cohort who do not actually have the disease; in turn, the health plans would be held responsible for providing disease-specific services to an enrollee for whom those services might be inappropriate. This would weaken the accuracy and utility of the resulting analysis. At the same time, we did not want to limit the criteria to the extent that we missed a large portion of the eligible population. The cohort definitions we tested did not prove satisfactory.

The depression analysis was further confounded by the fact that: 1) mental health services are carved out of the Maryland Medicaid program, 2) depression is a disease that is commonly treated by primary care providers, without guarantee of referral to the mental health system, and 3) the diagnosis of depression is subjective and can be used to describe a wide variety of presentations. Furthermore, our analysis relied solely on administrative data that included ICD-9 codes, but did not include more descriptive DSM-IV data. In developing our definition, we had to consider whether the characteristics of an enrollee with a depression diagnosis from a primary care provider were equivalent to the characteristics of an enrollee with a depression diagnosis from a mental health provider. For example, do enrollees who do not seek services in the mental health system have a milder form of the disease than those who do seek mental health services? In turn, would the performance expectations for these providers and enrollees be the same? Should we only evaluate enrollees who have accessed mental health services?

The challenges of defining a depression cohort are complex, but not insurmountable. This speaks directly to the need to be familiar with the data. We need to do additional work to better understand the coding patterns in the data and the utilization patterns of enrollees with depression in both the somatic and mental health system. We also need to be confident about our ability to reliably measure disease-specific services that are appropriate to the population. With additional time, we should be able to repeat our analysis on a cohort of enrollees with depression.