Project Report

Minnesota Health Care Risk Adjustment Methodology Development and Testing

Prepared for: Minnesota Department of Human Services

Submitted by: The Lewin Group

December 20, 2015

This document has been reviewed and confirmed to adhere to Accessibility Standards, Section 508 of the Rehabilitation Act, 29 U.S.C §79d.

This document was developed through funding provided by the grant program outlined under the Catalog of Federal Domestic Assistance Number 93.609 from the U.S. Department of Health and Human Services, Centers for Medicare & Medicaid Services. However, these contents do not necessarily represent the policy or views of the U.S. Department of Health and Human Services, and you should not assume endorsement by the Federal Government.

LEWINGROUP®

Consultant's Report

Table of Contents

Executive Summary

Overview

The Minnesota Department of Human Services (DHS) received grant funding from the U.S. Department of Health and Human Services, Centers for Medicare & Medicaid Services (CMS) to pursue the testing, collection, and reporting of the Initial Core Set of Health Care Quality Measures for Medicaid-eligible Adults. Within the scope of this funding opportunity, DHS sought to develop a risk adjustment methodology to enhance the use of the Initial Core Set of Health Care Quality Measures for Medicaid-Eligible Adults and enable more accurate comparison between managed care organizations (MCOs). DHS contracted with The Lewin Group to evaluate current health care risk adjustment methodologies and test usability by the Initial Core Set of Health Care Quality Measures for Medicaid-Eligible Adults.

Quality measurement is used to compare quality of care across areas such as health systems, providers, and payers. A risk of quality measurement is that it often fails to account for the differences in outcomes that arise when patients with different risk profiles disproportionally affect the performance of their providers. Risk adjustment seeks to account for patient characteristics that are outside of the provider's control, so that residual differences in performance more accurately reflect true quality differences. Risk adjustment accounts for patient-related attributes and allows measurement of health care quality to be comparable across providers and organizations seeing different mixes of patients. Risk adjustment may include adjustments for health status, socioeconomic factors, and other intrinsic patient factors.[1] The optimal risk adjustment system or technique to use may vary by quality measure, provider type, and type of data collected.

Phase One

Phase one of this project was an evaluation review that consisted of an environmental scan that identified appropriate methods and documented findings and recommendations for DHS. This evaluation review aimed to identify risk adjustment needs, resources, and current practices that may apply to the Initial Core Set of Health Care Quality Measures for Medicaid-Eligible Adults in managed care plans or fee-for-service arrangements. For each measure, risk adjustment examples from other states, CMS, commercial health plans, or other entities were considered, with emphasis on those that have been approved by the National Quality Forum (NQF) or Minnesota Community Measurement (MNCM).

The evaluation review first explored methodologies to risk adjust health care quality measures. Commercial "off-the-shelf" methodologies were analyzed alongside unique strategies devised by entities such as CMS, State Medicaid Agencies, the National Committee for Quality Assurance (NCQA), and commercial health plans to account for patient characteristics in the comparison of:

- The Plan All-Cause Readmission Rate Measure for Medicaid populations
- Medicare quality measures
- Minnesota Physician Clinic Quality Measures
- Pediatric quality indicators
- Health Employer Data and Information Set (HEDIS) and HEDIS Relative Resource Use Measures
- Consumer Assessment of Health Plans Study (CAHPS) Survey
- Home health agency quality measures

- Nursing home quality measures
- Hospital and surgery quality measures, and
- Agency for Healthcare Research and Quality (AHRQ) quality indicators.

Selecting a risk adjustment methodology requires the consideration of many factors, such as alignment with existing federal and state quality measurement initiatives, the availability of useful data or feasibility of data collection, achieving meaningful predictive performance, and avoiding unintended consequences in the process. CMS encouraged states to leverage existing methods and infrastructures for data collection and reporting, such as the Health Information Technology for Economic and Clinical Health (HITECH) and Medicaid Management Information Systems (MMIS). Minnesota administers the State Quality Reporting and Measurement System (SQRMS), through which physician practices may directly report quality performance information to the State. Additionally, CMS or other federal quality improvement activities, such as the National Quality Strategy, Strong Start Initiative, Partnership for Patients, and Million Hearts Initiative were studied for opportunities.

The criteria for selecting a risk adjustment methodology were presented to DHS, including: a section on the selection of risk factors, using socioeconomic risk factors, data sources, evaluating risk adjustment methodology performance, and maximizing the impact of risk adjustment. Lewin presented a preliminary recommendation of a risk adjustment methodology for DHS to use on the Initial Core Set of Health Care Quality Measures for Adults. The recommendation was for use of the Adjusted Clinical Groups (ACGs) model, with member stratification based on member health (i.e., acuity level) and socio-demographic characteristics.

Currently, the State employs the Johns Hopkins Adjusted Clinical Groups (ACG) system for premium rate setting for Managed Care Minnesota Health Care Programs (MHCP), for supporting patient risk stratification in the Patient Centered Health Care Home program, and plans to use it for total cost of care shared savings calculations in the Medicaid Health Care Delivery System Demonstration. While ACGs have historically been used for risk adjusting cost of care, the system is also designed to profile the performance of providers and health plans. Due to the level of market penetration in Minnesota, ACGs are a known commodity and accepted by the State's MCO stakeholders. The potential to use the ACG system for quality measurement risk adjustment and use of the Chronic Illness and Disability Payment System (CDPS) for disabled population was examined in this study.

Phase Two

Phase two of this project consisted of testing and implementing the risk adjustment methods. DHS grouped the measures into related clinical areas and selected 19 measures of interest. The measures encompassed the following measure groups: preventive women's health, chronic, mental health, behavioral, chronic hospitalization, and treatment measures. External research and DHS' interest guided the selection of characteristics to potentially include in the analysis. Characteristics of particular interest were patient-related that could influence the outcome on a given quality measure but were beyond the control of an MCO. Data from nine Minnesota MCOs and FFS data was provided for this analysis. The study employed robust statistical methods (e.g., logistic regression) to guide the selection of appropriate characteristics to include in the risk adjustment for each of the select adult quality measures. DHS' expertise regarding their population and policy considerations were additional factors used to guide the process. Statistical models were developed and applied to risk adjust the respective measures for each of DHS' contracted MCOs. These models adjust an MCO's rate for a given quality measure based on the characteristics identified as important for risk adjustment for that measure and an MCO's enrollment mix on those characteristics.

Lewin found that nearly all (17 of the 19) measures were suitable for risk adjustment. The outcomes were related to patient characteristics that differed across the MCOs. These characteristics are beyond the control of the MCOs and the distributions of these characteristics commonly differed across MCOs. This made risk adjustment suitable to account for these patient-related attributes and to facilitate more equitable comparisons across patient mixes.

The risk adjustment identified clinical and sociodemographic characteristics that were related to outcomes. The clinical factors encompassed a member's overall health risk, whether the member had a developmental disability, was enrolled in Medicaid due to a disability, was frail, had a mental health condition, or was identified as having a substance abuse issue. The sociodemographic characteristics encompassed a member's age, gender, education, language, race and ethnicity, and whether the member lived in a metropolitan county.

Health risk and age were consistently influential factors. The remaining clinical and sociodemographic characteristics had mixed results and showed a general pattern of smaller influence relative to health risk and age. Although, sociodemographic characteristics and health status can be correlated and prior adjustment for health could lessen the observable impact of these characteristics. Even so, a sociodemographic characteristic sometimes had larger influence on a select quality measure; this suggests sociodemographic characteristics should be considered when exploring risk adjustment of quality measures. The results also suggest that targeted clinical characteristics that might not be fully captured in a health risk measure could also be considered when exploring risk adjustment.

The influence of risk adjustment on the quality measures was also mixed. While, in general, there appeared to be a fairly even mix of risk adjustment increasing and decreasing rates, there was a slight tendency towards improved rates on the quality measures. Moreover, we observed that risk adjustment can have a different impact within subgroups of the MCOs' populations such as age groups. Therefore, this supports that risk adjustment results that include sociodemographic characteristics should be stratified to help ensure any potential disparities are not masked.

These analyses empirically demonstrate the influence of an array of clinical and sociodemographic characteristics on adherence rates among the Initial Core Set of Health Care Quality Measures for Medicaid-Eligible Adults. This is important because (1) MCOs commonly have different mixes of patients across the characteristics we found to influence these quality measures and (2) these patient-related attributes are beyond the control of MCOs. Consequently, not accounting for these patient-related differences could result in imperfect comparisons when contrasting groups such as MCOs. Ultimately, these analyses exhibit the need and appropriateness of risk adjusting quality measures when aiming to compare subgroups such as MCOs. Given the demonstrated relationships of the patient-related characteristics with the quality measures and that these characteristics differed across MCOs, this risk adjustment approach is recommended for Minnesota DHS to allow more accurate comparisons of MCOs. This approach is applicable to, and recommended for, other states with similar aims of comparing quality measure results across groups such as MCOs.

Summary and Evaluation

This project was successful and highly sustainable. The overall goals of the project were to make accurate comparisons of the quality of care provided by each MCO, to account for differences in the health status of the members enrolled in each MCO, and risk adjust the selected quality measures to account for the complex characteristics of the Medicaid population. Through the evaluation review, analyses, and delivery of the risk adjusted quality measures these goals were met. While the original goal was to explore risk adjusting only five to seven measures, seventeen measures were successfully risk adjusted.

Regarding sustainability, the ACGs are already in place and can be combined with existing enrollment data. Moreover, the SAS code that Lewin provided DHS was set up to ease replicability across additional measures and to ease refreshing results going forward. Lewin had a very positive experience working with DHS. The communication between DHS and Lewin created a cohesive partnership which allowed for iterative data analysis and an open discussion on what worked and what could be improved.

List of Acronyms

Accountable Care Organizations (ACO)
Adjusted Clinical Groups (ACG)
Affordable Care Act (ACA)
Aged, Blind, and Disabled (ADB)
Agency for Healthcare Research and Quality (AHRQ)
All Patient Refined Diagnosis Related Groups (APR-DRG)
American College of Surgeon's National Surgical Quality Improvement Program (ACS NSQIP)
Angiotensin Converting Enzyme (ACE)
Angiotensin Receptor Blockers (ARB)
California Coronary Outcomes Reporting Program (CCORP)
Centers for Medicare & Medicaid Services (CMS)
Chronic Illness and Disability Payment System (CDPS)
Chronic Obstructive Pulmonary Disease (COPD)
Confidence Interval (CI)
Consumer Assessment of Health Plans Study (CAHPS)
Coronary Artery Bypass Grafting (CABG)
Current Procedural Terminology (CPT)
Diagnostic Cost Group/Hierarchical Condition Category (DCG/HCC)
Episode Risk Grouper (ERG)
Evidence Based Medicine Symmetry Grouper (EBM)
Expanded Diagnosis Cluster (EDCs)
Fee-for-Service (FFS)
Generalized Estimating Equations (GEE)
Generalized Linear Mixed Modeling (GLMMIX)
Health Care Cost and Utilization Project (HCUP)
Health Care Finance Administration-DRG (HCFA-DRG)
Health Employer Data and Information Set (HEDIS)
Health Information Technology for Economic and Clinical Health (HITECH)
Health Outcomes Survey (HOS)
Healthcare Common Procedure Coding System (HCPCS)
Independent Review Entity (IRE)
Inpatient Hospital Stays (IHS)
Inpatient Quality Indicators (IQIs)
Inpatient Rehabilitation Facility (IRF)
International Classification of Disease (ICD)
Long-Term Care Hospital (LTCH)
Low Birth Weight (LBW)
Managed Care Minnesota Health Care Programs (MHCP)
Managed Care Organizations (MCO)
Medicaid Management Information Systems (MMIS)
Medicare Severity Diagnosis Related Groups (MS-DRG)
Minnesota Community Measurement (MNCM)
Minnesota Department of Human Services (DHS)
Minnesota Quality Incentive Payment System (QIPS)
National Committee for Quality Assurance (NCQA)

National Drug Code (NDC)
National Quality Forum (NQF)
Odds Ratio (OR)
Outcome and Assessment Information Set (OASIS)
Outcome-based Quality Improvement (OBQI)
Pay-for-performance (P4P)
Prevention Quality Indicators (PQI)
Receiver Operating Characteristic (ROC)
Relative Resource Use (RRU)
Resource Utilization Bands (RUBs)
Risk-adjusted Mortality Rate (RAMR)
Rural-Urban Continuum Codes (RUCC)
Social Security Income (SSI)
State Quality Reporting and Measurement System (SQRMS)
Temporary Aid for Needy Families (TANF)
U.S. Department of Health and Human Services (HHS)
Variance Inflation Factor (VIF)

Introduction

A. Background and history

The Minnesota Department of Human Services (DHS) received grant funding from the U.S. Department of Health and Human Services, Centers for Medicare & Medicaid Services (CMS) to pursue the testing, collection, and reporting of the Initial Core Set of Health Care Quality Measures for Medicaid-eligible Adults.

Within the scope of this funding opportunity, DHS seeks to develop a risk adjustment methodology to enhance the use of the Initial Core Set of Health Care Quality Measures for Medicaid-Eligible Adults and enable more accurate comparison between managed care organizations (MCOs) and accountable care organizations (ACOs). Through risk adjustment DHS aims to account for patient characteristics that are outside of the provider's control so that residual differences in performance reflect true quality differences. That is, risk adjustment accounts for patient-related attributes and allows measurement of health care quality to be comparable across providers and organizations seeing different mixes of patients. DHS contracted with The Lewin Group to evaluate current health care risk adjustment methodologies and test for appropriateness and feasibility of use by the Initial Core Set of Health Care Quality Measures for Medicaid-Eligible Adults. Phase one of this project was an evaluation review that consisted of an environmental scan that identified appropriate methods and documented findings and recommendations for DHS. Phase two of this project consisted of testing and implementing the risk adjustment methods.

B. Health Care Quality and Performance Measures

Quality measures generally fall into two main categories: outcome measures and process of care measures. Outcome measures assess how patients fare under a provider's care, while process of care measures describe what is done to and for patients in the care setting.[2] Risk adjustment is particularly important for outcome measures because patient outcomes may be as much driven by patients' physical and socio-demographic characteristics, over which clinicians have little control, as by quality of care.[3] However, risk adjusting outcomes of routine outpatient care involving chronic conditions is cited as challenging due to complexities in monitoring patient risk factors over an extended period. The Plan All-Cause Readmission Measure is an example of an outcome measure within the Medicaid Adult Core Set. Several different risk adjustment methodologies are currently being used by states to adjust this measure. Two states are using a CMS-developed methodology intended for Medicare and commercial populations, which identifies risk adjustment categories and weights for each Index Hospital Stay (IHS) based on a number of clinical and basic demographic risk factors.

Process measures often do not require full risk adjustment, and may benefit from risk stratification or by applying specific eligibility criteria and exclusion criteria for instances in which the recommended care may not be applicable.[4] For example, CMS reported separate measures for High Risk and Low Risk Patients for the Nursing Home Compare System's Long-Stay Residents quality measures. However, process measures may also benefit from risk adjustment, particularly those that depend on patient adherence (such as medication adherence and obtaining preventative screenings and immunizations).[5] These measures may also require adjustment for factors beyond patients' clinical attributes that may otherwise confound performance measurement. However, such elements may be difficult to measure or incorporate into the risk adjustment systems because they often touch upon personal motivation, educational attainment, financial resources, transportation, time off work, preferences for care, etc.[6] On the other hand, process measures often specify risk factors common among patients who seek a particular type of service, which can serve as inclusion or exclusion criteria for subsets of patients during risk adjustment.

When performing risk adjustment, it is necessary to take into account that certain risk factors may be not be measurable or consistently reported. Furthermore, literature suggests that risk adjustment may unintentionally thwart care improvement initiatives in certain pay-for-performance settings if providers are rewarded (or not penalized) for providing unexceptional or second rate care to patients who have higher socio-demographic risk factors that were accounted for in the risk adjustment process.[5]

Additional information about quality measurement is available in Appendix A.

C. Risk-Adjusting Quality Measures

What is risk adjustment?

Health-based risk adjustment is a process traditionally used to adjust capitation payments for managed care organizations (MCOs) and provider groups to account for differences in cost for treating beneficiaries with different conditions. Diagnostic information from fee-for-service claims or encounter data is used to develop a profile of the beneficiary's condition panel to predict future health care costs. Additionally, beneficiary characteristics such as age, gender, health status, and geographic area are taken into account to identify beneficiaries expected to have higher health care costs.

Risk adjustment of health care quality measures is an extension upon the traditional use of risk adjustment for payment and involves using statistical methods to adjust for patient-related factors. This allows for more accurate assessment of quality of care and more equitable comparison between MCOs, health care facilities, or providers. Without risk adjustment, health plans and providers may find incentive to enroll healthier patients avoid sicker patients. When risk adjustment is performed correctly, it should reduce the incentives for providers to avoid patients who could adversely impact their performance outcomes.

The National Quality Forum (NQF) currently recommends risk adjustment for outcome measures for clinical factors, on the basis that patients who have numerous or severe conditions are inherently likely to have worse outcomes, regardless of the quality of care provided.[7] When sociodemographic characteristics are included in the risk adjustment of quality measures, the NQF also recommends stratifying results.[8] Stratification refers to examining the quality measures separately for each subgroup and helps identify potential health disparities.

Additional information about risk adjustment is available in Appendix A.

Why apply risk adjustment to health care quality measures?

Government health care agencies, MCOs, health care facilities, providers, and patients are increasingly in search of objective measures of outcome and other clinical performance measures. To that end, risk adjustment is becoming an increasingly important tool for making clinical, administrative, and economic decisions due to its ability to isolate outcomes of treatment interventions from inherent patient characteristics and risk factors.[9] While many risk factors greatly impact patient outcomes, they typically do not affect the quality of care provided.

The purpose of risk adjustment centers on accounting for patient characteristics that influence outcomes because different groups of people can have different characteristics (e.g., patient mix). For example, people with greater risk factors might select plans that offer specific benefits that better meet their needs, which could result in biased risk pools.[10] In the absence of risk adjustment, performance measures that assess quality of care on the basis of patient outcomes may be biased in favor of plans or providers with an inherently healthier patient panel. In debating whether or not to apply risk adjustment to

sociodemographic factors, NQF considered providers who avoid serving disadvantaged populations to evade being labeled as a poor performer, thereby worsening access to care for vulnerable beneficiaries.

The Affordable Care Act (ACA) prohibits health plans in the individual and small group markets from denying coverage based on pre-existing conditions or health status. To this end, risk adjustment is needed to ensure that plans will not enroll a disproportionate share of healthier patients. The ACA also includes provisions to increase value-based purchasing and accurate public reporting of quality of care to promote improved care quality and efficiency, both of which are bolstered by risk adjustment.

Who might benefit from risk adjustment of quality measures?

Risk adjustment may alleviate certain issues associated with health care delivery, access to care, and performance measurement.

Patients and consumers

Risk adjustment may improve access to quality health care for more beneficiaries. Risk adjustment accounts for patient characteristics that influence outcomes when different groups of people have different characteristics (e.g., patient mix). For example, people with greater risk factors might select plans that offer specific benefits that meet their needs, which could result in biased risk pools. Patients and their families will benefit from a clearer, apples-to-apples comparison of providers' performance, which could enhance decision-making and create a more informed patient base.

MCOs and providers

Risk adjustment allows MCOs to enroll, or providers to treat, complex patients without fear of being "penalized" when certain performance metrics are compared to those who enroll or treat relatively healthier patients. In the context of performance-based incentive programs, such as pay-for-performance (P4P), risk adjustment provides a way to accurately account for inherent differences in patient panels, allowing for an unbiased comparison of provider performance.

Additionally, risk adjustment is essential in helping MCOs and providers with internal quality improvement initiatives and performance measurement activities by allowing them to compare results with peer entities. Comparison of results that are not risk adjusted may be misleading. Risk adjustment also helps internal quality improvement efforts by tracking quality outcomes over time and establishing a baseline adjusted for patient characteristics and risk factors of any given time frame.[9]

State Medicaid Agencies

Increasingly, state Medicaid agencies are implementing initiatives focused on comparing provider performance and those performance measures can be used to inform P4P. Risk adjustment supports equitable comparison of clinics, medical groups, MCOs and hospital performance. Risk adjustment of quality measures increases accountability for performance and public awareness for differences in the quality of care provided by different entities. It has the potential to improve the comparability of quality metrics both across providers and over time. It is instrumental for incentive-based performance incentive programs and provides an accurate baseline for assessing quality of care provided within states, and in comparison to other states.

How is risk adjustment applied to quality measures?

The development of risk adjustment methods is largely dependent on three primary considerations: 1) the statistical method, 2) choice of risk factors, and 3) selection of data sources. Depending on the particular outcome measure, care environment, and purpose, risk adjustment methodologies differ widely in terms of their risk factor specifications, weighting schemes, and application.[2]

Risk adjustment of quality measures may be carried out in many ways. When it is not feasible to adopt indicators that have already been risk adjusted by an intermediary, such as CMS, organizations may choose to use "off-the-shelf" risk adjustment models that are built into many software programs. For example, this category includes ACGs, the Chronic Disability & Illness Payment System (CDPS) for the Medicaid population, and several others. Alternatively, agencies may develop their own risk adjustment models for selected measures or conditions. Often, customized modeling is ideal when specific strengths are present in the data, or when existing methodologies are not readily available.

Risk adjustment of quality measures is performed across agencies, health care facilities, and payers. The range of commonly used techniques varies by computational intensity, and may entail:[11]

- Direct stratification of outcome results for patients in different risk categories based on severity of illness, income, and other factors. Providers receive a performance score for each risk category, and comparisons are made across categories rather than based on an overall total score.

  For example, direct stratification has been used to compare differences in mortality rates for cardiovascular and non-cardiovascular deaths in patients starting dialysis to that in the general population.[12]

- Comparing observed to expected outcomes by applying indirect standardization where the expected outcomes are determined by applying stratum-specific rates. Indirect standardization is based on identifying various categories that could be examined by population and provider. Indirect stratification adjusts for variations in frequencies of specific risk factors in two study groups to enable comparison of the expected rates of a particular outcome.

  For example, indirect risk adjustment has been used to compare the pre-38 week gestation risk of cesarean delivery in two groups (women treated by family physicians compared to those treated by obstetrician-specialists).[13]

- Organizational stratification to create peer groups or sub-organizations with similar patient case mix. Organizations may be ranked in quintiles or deciles based on the proportion of patients below a certain threshold of the federal poverty level.

  For example, ACGs were used to risk adjust HealthPartners' Total Cost of Care and Total Resource use measures (which assesses the frequency and intensity of services utilized to manage a provider's patients) by adjusting a payer's membership for variations in patient disease burden.[14]

- A combination of a statistical risk model and stratification, which may entail statistical adjustment for clinical factors and stratification for sociodemographic factors, or applying different statistical models for each stratum that are then used to form an overall performance score.

  Risk stratification entails computing performance scores separately by selected variables (i.e., strata) and is commonly applied before using a statistical risk adjustment model.

- A limited risk adjustment strategy can be applied through exclusion criteria, intended to create homogenous comparison groups.

    The basic step in applying exclusion criteria is defining the relevant population (denominator) based on outcome-specific criteria.

Impact of risk adjustment on clinical practice and quality of care

Risk adjustment has had a significant positive impact on clinical practice and research for its roles in helping to isolate the outcomes of treatment interventions from research subjects' inherent pre-existing conditions and risk factors.[9] Risk adjustment helps clinicians make more informed assessments of new research findings and how best to apply those research findings to their patients. Additionally, an understanding of risk adjustment allows clinicians to have a better grasp of benchmarking, performance measurement, and quality improvement standards.[9]

Phase One – Evaluation Review

Phase one of this project was an evaluation review that consisted of an environmental scan that identified appropriate risk adjustment methods and documented findings and recommendations for DHS.

Selecting a risk adjustment methodology requires the consideration of many factors, such as alignment with existing federal and state quality measurement initiatives, the availability of useful data or feasibility of data collection, achieving meaningful predictive performance, and avoiding the encouragement of inherent perverse incentives in the process. CMS encourages states to leverage existing methods and infrastructures for data collection and reporting, such as the Health Information Technology for Economic and Clinical Health (HITECH) and Medicaid Management Information Systems (MMIS). Minnesota administers the State Quality Reporting and Measurement System (SQRMS), through which physician practices may directly report quality performance information to the State. Additionally, CMS or other federal quality improvement activities, such as the National Quality Strategy, Strong Start Initiative, Partnership for Patients, and Million Hearts Initiative should be studied for opportunities to harmonize with the task at hand.

Currently, the State employs the Johns Hopkins Adjusted Clinical Groups (ACG) system for premium rate setting for Managed Care Minnesota Health Care Programs (MHCP), for supporting patient risk stratification in the Patient Centered Health Care Home program, and plans to use it for total cost of care shared savings calculation in the Medicaid Health Care Delivery System Demonstration. While ACGs have historically been used for risk adjusting cost of care, the system is also designed to profile the performance of providers and health plans. The potential to use the ACG system for quality measurement risk adjustment and use of the CDPS system for disabled population was examined.

This evaluation review aimed to identify risk adjustment needs, resources, and current practices that may apply to the Initial Core Set of Health Care Quality Measures for Medicaid-Eligible Adults in managed care plans or fee-for-service arrangements. For each measure, risk adjustment examples from other states, CMS, commercial health plans, or other entities were considered, with emphasis on those that have been approved by National Quality Forum (NQF) or Minnesota Community Measurement (MNCM).

Initial Core Set of Health Care Quality Measures for Adults Enrolled in Medicaid (Medicaid Adult Core Set)

CMS and the Agency for HealthCare Research and Quality (AHRQ) facilitated the identification and prioritization of an initial core set of 26 adult health care quality measures for voluntary use by states.[15] As of 2013, the Plan All-Cause Readmission measure is the only measure in the Medicaid Adult Core set for which risk adjustment is being performed. However, this measure currently does not have a specific risk adjustor for the Medicaid population.

Table 1 presents the Medicaid Adult Core Set of quality measures, including the measure steward for each measure, identification of measures collected and reported by DHS in 2013 and those planned for production in 2014. The measure steward is responsible for the maintenance of the measure or measure set. Maintenance includes updating measures as new clinical evidence is available and updating the codes that are tied to certain technical specifications. Measure 27: HIV Viral Load Suppression is new for 2014 and is scheduled to replace Measure 16: Annual HIV/AIDS Medical Visit.

Table 1. Initial Core Set of Health Care Quality Measures for Adults Enrolled in Medicaid.

Measure Name	Measure Steward	2013	2014
1. Flu Shots for Adults (50-64)	NCQA/HEDIS	Yes	Yes
2. Adult BMI	NCQA/HEDIS	No	No
3. Breast Cancer Screening	NCQA/HEDIS	Yes	Yes
4. Cervical Cancer Screening	NCQA/HEDIS	Yes	Yes
5. Medical Assistance with Tobacco Use Cessation	NCQA/HEDIS	Yes	Yes
6. Screening for Clinical Depression and Follow Up	CMS	No	No
7. Plan All-Cause Readmission	NCQA/HEDIS	No	No
8. PQI 01:Diabetes, Short-term Complication Admission Rate	AHRQ	Yes	Yes
9. PQI 05: Chronic Obstructive Pulmonary Disease Admission Rate	AHRQ	Yes	Yes
10. PQI 08: Congestive Heart Failure Admission Rate	AHRQ	Yes	Yes
11. PQI 15: Asthma in Younger Adults Admission Rate	AHRQ	Yes	Yes
12. Chlamydia Screening in Women Aged 21 -24	NCQA/HEDIS	Yes	Yes
13. Follow-Up After Hospitalization for Mental Illness	NCQA/HEDIS	Yes	Yes
14. PC – 01: Elective Delivery	The Joint Commission	No	No
15. PC – 03: Antenatal Steroids	The Joint Commission	No	No
16. Annual HIV/AIDS Medial Visit	NCQA	No	Retired
17. Controlling High Blood Pressure	NCQA/HEDIS	No	No
18. Comprehensive Diabetes Care: LDL-C Screening	NCQA/HEDIS	Yes	Yes

Measure Name	Measure Steward	2013	2014
19. Comprehensive Diabetes Care: Hemoglobin A1c Testing	NCQA/HEDIS	Yes	Yes
20. Antidepressant Medication Management	NCQA/HEDIS	Yes	Yes
21. Adherence to Antipsychotics for Individuals with Schizophrenia	NCQA/HEDIS	No	Yes
22. Annual Monitoring for Patients on Persistent Medications	NCQA/HEDIS	No	Yes
23. CAHPS Health Plan Survey 5.0H – Adult Questionnaire	AHRQ NCQA/HEDIS	Yes	Yes
24. Care Transition – Transition Record Transmitted to Health Care Professional	AMA/PCPI	No	No
25. Initiation and Engagement of Alcohol and Other Drug Dependence Treatment	NCQA/HEDIS	Yes	Yes
26. Postpartum Care Rate	NCQA/HEDIS	Yes	Yes
27. HIV Viral Load Suppression	HRSA	NA	No

Risk Adjustment Models

Risk adjustment serves three primary functions:[16]

- Adjust capitation payments for expected future expenditures based on member health status
- Create risk profiles or identify candidates for disease management
- Adjust for observed differences in performance measures, utilization, and or/cost based on differences in patient characteristics and acuity

While contemporary models were designed for payment, many are capable of effectively performing risk adjustment to enable comparison between providers for performance measurement.

Patented risk adjustment models generally follow one of two basic structures: categorical models (or "groupers") or additive models. Categorical models use information about each case to assign it to one of several categories that match its level of risk for a particular outcome. An example is the ACG system, which groups each person into one of many categories based on a morbidity profile, age, and gender. The system then calculates a prediction for each person based on the average value of the outcome of interest for all persons in the same group. Additive models, such as the Diagnostic Cost Group (DCG), Hierarchical Condition Category (DCG/HCC) model and the CDPS model, makes predictions directly from patient profiles by first assigning each case a score based on the sum of the coefficients for each chronic condition included in the model.

Additive models compute a risk score for each member that is an estimate of their future or current health care costs. Additive models assign members to mutually exclusive homogenous risk groups based upon their expected health care utilization. The risk score generated by additive models is a predictor of expenditures, not expected adherence to a quality measure. Therefore it is not directly applicable to risk adjusting quality measures. Categorical models group members with comparable health care needs. Adherence rates can be measured for each risk group, and quality measures for a provider can be risk adjusted based upon their distribution of members across the risk groups. In order to utilize an additive model to risk adjust quality measures, members should first be assigned to mutually exclusive risk groups based upon their risk scores and other measures of risk generated by the model. So both model structures can be utilized, however additive models require an additional step prior to their application to risk adjust quality measures.

Another primary consideration in methodology selection is data availability and quality. For managed care programs, several states have experienced difficulty in collecting accurate and complete diagnostic data from their managed care plans. Pharmacy data is generally the most complete encounter data source and the pharmacy based groupers may provide more accurate measures of risk during the start-up phase of a new managed care program. However, if diagnostic data quality is an issue then the quality measures that utilize this data may also be inaccurate.

Stratification

Stratification techniques are useful for controlling for the confounding effects of categorical risk factors when individual strata are meaningfully defined and homogenous. However, stratification may only be used to adjust for several risk factors simultaneously, and can only be used to compare the association between one risk factor and one outcome variable at the same time.

In direct stratification, a standard population is identified and the incidence rate is calculated for each stratum of the potentially confounding variable. Next, the expected number of outcomes in each stratum of the standard population is computed by multiplying by stratum-specific rates to the number of subjects

in the standard population. The standardized outcome rates are then calculated by dividing the overall expected outcomes in the standardized population by the total number of individuals in the standard population.

For indirect stratification, the ratio of the total number of observed outcomes to the number of expected outcomes serves as an approximation for the risk factor-adjusted relative risk ratio. Indirect standardization is particularly useful when population groups are small.

Logistic regression

Logistic regression is a widely used methodology for adjusting discrete outcome measures in health care. It allows for the inclusion of a large quantity of risk factors, which is an advantage above other risk adjustment techniques such as stratification or standardization. Logistic regression results have been shown to be empirically equal or superior to results produced using other risk adjustment techniques, in terms of explanatory power and clinical and statistical utility and understandability.[17] The estimated parameters, associated statistics, goodness of fit and diagnostic methods for testing are well established.

The c-statistic is the most prevalent summary measure used to evaluate logistic regression models, and it assesses the model's ability to discriminate between those who exhibit the outcome, and those who do not. It is the proportion of pairs in the dichotomy in which the predicted probability of death (or other dichotomous outcome) is higher for the person who died.[18]

The predictive model is used to obtain a predicted outcome as a function of a variety of risk factors for patients attributed to a provider. These outcomes are then averaged to determine a provider-level expected outcome rate.[19]

Hierarchical modeling

While logistic regression recognizes random variation of patients within providers, it does not account for random variation among providers, which tends to increase standard error. Logistic regression also does not adjust for the clustering of patients within providers, which results in dependencies between outcomes within the same cluster.[19] Hierarchical models (also termed multilevel or random effects models) adjust for these issues by providing a framework for integrating variation at different risk levels. "Nested" data occurs when data are generated in groups at the physician, practice group, or regional levels, with each level requiring a different set of independent variables. There is a growing body of literature on the use of hierarchical models for multilayer or nested data structures in the assessment of provider performance.

A. "Off-the-shelf" risk adjustment models

There are numerous widely-used risk adjustment models, each with a distinct methodology. Although developed for different uses or populations, they have largely become interchangeable over time. The ACGs, CDPS, DRGs, DCGs, and HCCs are grouper models that apply pre-defined algorithms to identify conditions present within a given population. These models use diagnosis data from administrative records to identify member conditions and score them for relative risk in order to predict health care costs. Models are calibrated by using source data to periodically update the model to reflect changes such as new medical technologies, new prescription drugs, changes in medical practice patterns, or changes in provider coding practices. Models can also be re-calibrated based local conditions by developing risk weights through a linear regression model based on local data.

Although risk adjustment strategies may be developed based on available data, there are three primary advantages of using "off-the-shelf" risk adjustment models:[20]

- Developing a risk adjustment model is a resource-intensive endeavor, particularly when a broad spectrum of conditions is to be included. Adding additional types of data, such as laboratory or clinical data, would also significantly increase the complexity of the development;
- Models must be continuously maintained to accommodate new codes. New drug codes, for instance, are released monthly; and
- Results from widely used risk adjustment models are easy to review and validate. The Society of Actuaries monitors the accuracy and usefulness of risk adjustment models and publishes comparative studies periodically.

Medical grouper models

Adjusted Clinical Groups (ACGs)

- **Developer:** Johns Hopkins University
- **Population:** General
- **Data:** Claims data, emphasis on demographic and diagnoses data; optional inputs include: pharmacy data, prior cost experience, utilization measures, and procedure information
- **Algorithm:** Combines diagnoses into mutually exclusive groups based on clinical information, resources used, and patient characteristics. The model follows an actuarial cell structure, based on Aggregated Diagnosis Groups (ADGs).

The ACG model is a diagnosis-based case-mix adjustment model for ambulatory populations that was developed in response to CMS's classification system for inpatient care, DRGs.[20] The system is based on components called Aggregated Diagnosis Groups (ADGs), which are groupings of diagnosis codes by severity and likelihood of persistence of the health condition over time. Individuals are assigned to one or more ADGs based on five clinical dimensions: duration, severity, diagnostic certainty, etiology, and specialty care. As individuals develop more conditions over time, their pattern of morbidities allows the ACG system to assign them to a single cell, which is aimed to capture clustering of comorbidities experienced by individuals over time.[20] This is particularly useful in identifying population sub-groups that are particularly vulnerable and are in immediate need of resources.

In addition to ADG categories, the model uses disease-specific Expanded Diagnosis Clusters (EDCs) and diagnostic indicators of probability of future hospitalization and medical frailty. The ACG system can also create pharmacy-based and combination models to assist in identifying high risk individuals for case management. Model updates in 2010 have also incorporated additional data inputs, such as procedure codes and provider information. The model can be calibrated to reflect local conditions.

The "Cost of care: total resource use population-based member per month (PMPM) index" is one of the HealthPartners' population-based Total Cost of Care and Total Resource Use measures that is used to assess total resource use and cost.[i] This measure may be used by health plans and providers to help

[i] The Resource Use Index (RUI) is a risk-adjusted measure of the frequency and intensity of utilized services to manage patients, including all professional, facility inpatient and outpatient, pharmacy, lab, radiology, ancillary and behavioral health services.

identify areas where cost may be lowered by improving resource efficiency or using less expensive but equally effective resources, or to identify overuse and underuse of certain health care services.[28]

Risk adjustment is used because the total cost of care and resource use can vary between providers based on the cost per member and/or resource use per member, all other factors equal. The measure uses ACGs to adjust for variations in disease burden of patients and institute peer grouping controls for patient demographics, provider types, and attributes of the insurance product. Theoretically, the remaining factors reflect what the provider can directly influence.

Chronic Illness and Disability Payment System (CDPS)

- **Developer:** University of California, San Diego
- **Population:** Medicaid, with separate weights for adult and child populations
- **Data:** Claims data; emphasis on demographic information, diagnoses, length of enrollment, dates of service, type of provider, procedural information, costs, and outpatient data
- **Algorithm:** Acuity categories are based on resource use, while combinations of diagnoses are based on clinical information and resource data. The model uses over 700 diagnosis groups combined into over 50 diagnosis subcategories. Patients may be assigned to multiple diagnosis groups.

CDPS is a diagnostic classification system originally developed to help state Medicaid programs make adjusted capitated payments for Medicaid beneficiaries, including Temporary Aid for Needy Families (TANF) and disabled Medicaid beneficiaries enrolled in either Supplemental Security Income (SSI). It was later also adapted for use in adjusting capitated Medicare payments to health plans under a different set of weights. Groups of diagnoses categorized by ICD-9 codes are further subcategorized by severity level. Weights are calculated for each group.

The model for Medicaid includes 20 major categories of diagnosis, and assigns members to at least one of 67 medical condition categories based on diagnosis codes and one of 16 age/gender categories. The medical condition categories are further divided into subcategories based on the degree of increased expenditures associated with the diagnosis. Based on a member's assigned medical condition category and age/gender group, and two sets of weights (one calibrated for the TANF population and another calibrated for the disabled population),the model predicts total medical costs for each member.

Diagnostic Cost Groups (DCG)

- **Developer:** Yale University, refined by 3M
- **Population:** General
- **Data:** Diagnoses, demographics, co-morbidities, acuity, severity of illness, prognosis, risk of death, resource intensity
- **Algorithm:** Diagnosis categories are hierarchically combined into Hierarchical Condition Categories. The model is not discharge-specific and focuses on resource consumption

The DCG model is a diagnosis-based model that is customizable depending on the population of interest (Medicaid, Medicare or commercial), source of data (inpatient only or all encounters), and purpose (payment or explanation). Diagnosis codes are grouped into clinically homogenous groups, called DxGroups, which are further assigned into 184 hierarchical condition categories. The DCG model also assigns each patient to one of 32 age/gender categories. The model predicts total medical cost for each patient based on the hierarchical condition categories and the age/gender category.

Hierarchical Condition Categories (HCC)

- **Developer:** CMS
- **Population:** Medicare
- **Data:** Diagnoses, demographics, co-morbidities, acuity, severity of illness, prognosis, risk of death, resource intensity
- **Algorithm:** Uses data to prospective estimate predicted costs for enrolled members during next year of coverage. Diagnosis categories are hierarchically combined into Hierarchical Condition Categories. The model is not discharge-specific and focuses on resource consumption

The HCC model is an additive DCG model developed for Medicare HMOs, focusing on resource consumption and total cost of care. CMS initially used a DCG/HCC combination model, which required encounter data from MCOs on physician office and hospital outpatient settings. After receiving complaints from MCOs about the burden of reporting encounter data, CMS designed a simplified version of the DCG model, called the CMS-HCC model. The model is not discharge-specific and does not incorporate all diagnoses because it is focused on adjusting for risk associated with select high-cost diagnoses.[28] The most recent model (CMS-HCCs) uses 70 HCCs used to determine reimbursement, selected from the original DCG/HCC models' 101 HCCs. Several attributes of the newest model include covering a broader range of health disorders and conditions, having well-defined diagnostic criteria, including conditions with significant expected health expenditures, and excluding highly discretionary diagnoses.

Drug -based models

MedicaidRx

- **Developer:** University of California, San Diego
- **Population:** Medicaid, with separate weights for adult and child populations
- **Data:** Prescription drug claims
- **Algorithm:** Prescription drugs are grouped into medical condition categories. Costs are predicted based on the medical condition category and age/gender category

The MedicaidRx model uses Medicaid pharmacy data to assign each individual to one or more of 45 medical condition categories based on the prescription drugs he or she uses, and also to one of 11 age/gender category.[21] Based on a combination of the medical conditions and age/gender category, the model predicts the overall medical costs for each member.

Pharmacy Risk Groups (PRGs)

- **Developer:** Ingenix
- **Population:** General, with focus on large managed care populations
- **Data:** Prescription drug claims
- **Algorithm:** Prescription drugs are first combined, then grouped into diagnostic categories. Patients may be assigned to multiple diagnostic categories. Model is calibrated by enrollment period

The PRG model uses prescription data and the Symmetry Episode Treatment Groups (ETGs) illness classification and episode bundling system to create markers of health risk that may indicate a patient's

disease prevalence, severity, and comorbidities.[22] PRGs support predictive modeling by identifying members at highest risk for high utilization and allows for more efficient disease and care management by stratifying members within particular diseases. It is particular beneficial for organizations that do not have access to complete medical claims data.

Episode groupers

Symmetry Episode Risk Groups (ERGs)

- **Developer:** Ingenix
- **Population:** General
- **Data:** Claims data and prescription drug claims
- **Algorithm:** Health care services are assigned to treatment groups, which are then categorized into an Episode Risk Group. This creates a clinical risk profile, from which a risk score can be generated by combining weighted ERG scores and demographic variables for each patient

The ERG model predicts current and future health care usage for groups and individuals using individual risk measures based on episodes of care methodology, medical and pharmacy claims information, and demographic variables to predict health risk.[22] The ERG risk scores are based on risk-adjusted episodes of care created by the Episode Treatment Groups (ETG) grouper. Each member is assigned to one of more than 120 medical risk groups, or "episode risk groups," based on diagnostic and procedural information from medical and pharmacy claims. Instead of relying on diagnoses data from individual medical encounter data, the ERG system uses episodes of care as markers of risk by applying the ETG classification and episode bundling system. Specific medical services are assigned episodes of care, and ETGs are able to prioritize related medical conditions. This enables easy identification of the episodes of care that are contributing most to a patient's risk.

B. Comparison of risk adjustment tools

Risk adjustment models differ in the health care data information they require and their approach to applying this information. For example, similar to CDPS and HCCs, ACGs assigns diagnosis codes to various categories, which are then used to predict expenditures based on diagnostic data. However, the ACG method of categorizing diagnoses differs significantly from the other two approaches because it assigns ICD-9-CM codes to diagnostic categories called ADGs based on expectations about a condition's effect on health; cost and likelihood on persistence, disability, reduced life expectancy; and need for diagnostic, specialist, therapeutic, and hospital care.[23] A comparison of select risk adjustment tools is presented in Table 2.

Table 2. Summary of Select Grouper Models.

Risk adjustment Models	Developer	Calibration Population	Data Requirements	Model Structure
Adjusted Clinical Groups (ACG)	Johns Hopkins University	General; local calibration	Demographics, diagnoses; Optional inputs include: pharmacy, prior cost experience, utilization measures, and procedures	Actuarial cell, predictive modeling suite (including diagnoses, pharmacy, and combination models)
Chronic disability payment system(CDPS)	University of California, San Diego	Medicaid and Medicare populations, with emphasis on disabled and TANF populations and separate weighting for adults and children	Demographics, diagnoses, length of enrollment, dates of services, type of provider, procedural information, category of service, costs, outpatient data	Model uses over 700 diagnostic groups, combined into over 50 diagnostic subcategories; model focuses on resource consumption
Diagnostic-Related Groups (DRG)	Yale University, refined by 3M	Inpatient Medicare population	Diagnoses, age, sex, co-morbidities, severity of illness, risk of dying, prognosis, treatment difficulty, need for intervention, resource intensity	Divides cases into 23 groups for Major Diagnostic Categories (MDC); within each MDC, patients are further divided into clusters
Diagnostic Cost Groups/Hierarchical Condition Categories (DCG/HCC)	RTI International and Boston University, with CMS funding	General	Demographics, diagnoses, costs	781 DxGroups, aggregated into condition categories; 101 hierarchical groupings of condition categories; models are not discharge-specific and focuses on resource consumption
MedicaidRx	University of California, San Diego	Medicaid, with separate weighting for adults and children	Pharmacy, age, sex	Model predicts overall medical costs for each member based on medical condition categories that he/she is assigned to (based on prescription drug usage) and age/sex category

Risk adjustment Models	Developer	Calibration Population	Data Requirements	Model Structure
Pharmacy Risk Groups (PRGs)	Ingenix	Large managed care populations, calibrated by enrollment period	Pharmacy claims	Categories of prescription drugs are mapped to diagnostic categories
Episode Risk Groups (ERGs)	Ingenix	General	Medical and pharmacy claims, demographic variables	All treatment information used in episode definition; An individual's health care services are assigned to treatment groups; each ETG is categorized into an ERG, creating a clinical and risk profile; a risk score is generated by combining weighted ERG scores and demographic variables for each person

Source: Adapted from Duncan IG, 2011. Healthcare Risk Adjustment and Predictive Modeling. ACTEX Publications.

Alternatively, some organizations choose to develop their own risk-adjustment models for selected measures, conditions, or procedures. This type of customized modeling can be ideal when there are particular strengths or features in the local data that would improve risk adjustment. For example, a "present on admission" coding is performed for every diagnosis in California, a data element that dramatically improves risk adjustment for the state's coronary artery bypass mortality rate measure.[24]

Certain quality measures may not require risk adjustment. Two alternative approaches to risk adjustment, commonly used for process measures, are exclusion and risk stratification. Exclusion entails excluding patients who do not qualify for the process of care in question from the denominator. The treatment or therapy may have not been successfully shown to provide a clear benefit to the patient or has proven to be medically inadvisable to administer to the patient. Alternatively, in risk stratification, patients are divided into two or more groups according to their expected risk. For example, health plans may be asked to report on the outcomes of their healthy and sick members separately, or provide separate analyses for members with different benefit designs. Stratification may be particularly useful for uncovering disparities in care and for rewarding health plans and physician groups that reduce such disparities.[25] However, the Agency for Healthcare Research and Quality (AHRQ) cautions that reporting stratified data may require larger sample sizes than reporting aggregated data.[26]

Evidence from the Literature

This section provides an environmental scan of past and current initiatives by Medicaid, Medicare, commercial payers, the National Committee for Quality Assurance (NCQA), home health agencies, nursing homes, hospitals, and other entities to risk adjust quality measures. For each risk adjustment methodology, relevant background is provided regarding the rationale for risk adjustment; the current status of the methodology; and developments to refine the methodology.

A. Risk adjustment for Plan All-Cause Readmission Rate Measure

Background

The Centers for Medicare and Medicaid Services has funded 26 states to develop the capacity to collect and report on a minimum of 15 measures from the Initial Set of Health Care Quality Measures for Medicaid-Enrolled Adults. CMS is seeking risk adjustment on one of the measures in the Medicaid Adult Core Set, the All-Cause Readmission measure, which computes the number of acute inpatient stays that were followed by an acute readmission for any diagnosis within 30 days, as well as the predicated probability of such an acute readmission. This measure does not currently have a Medicaid-specific risk adjustor, and states that report on this measure are advised to either describe the risk adjustment methodology used or postpone risk adjustment until a Medicaid-specific strategy is developed or approved by CMS.[27]

Data is currently reported for three metrics:

- Count of Index Hospital Stays (IHS) (denominator)
- Count of 30-Day Readmissions (numerator)
- Average Adjusted Probability of Readmission (rate)

Conceptual framework

A critical transition point in care occurs when a patient is discharged from the hospital. Clinical evidence suggests that inadequate care coordination or care transition processes at discharge can contribute to higher rates of morbidity and lead to rehospitalization and may be indicative of poor patient planning processes at the care facility.[28] The Plan All-Cause Readmission Rate Measure is intended to distinguish avoidable readmissions due to complications stemming from poor care coordination at discharge to rehospitalizations due to pre-existing comorbidities.[28]

Risk adjustment variables

According to the National Quality Measures Clearinghouse, risk adjustment weights should be calculated based on presence of surgeries, discharge condition, comorbidity, age, and gender for each index hospital stay.[28]

Risk adjustment methodologies among commercial payers

Certain commercial payers are also risk adjusting the All-Cause Readmission Measure. For example, PacifiCare's methodology begins by applying the CMS DRG Grouper, All Patient Refined Diagnosis Related Groups (APR-DRG) Grouper, or Medicare Severity Diagnosis Related Groups (MS-DRG) grouper to inpatient claims data. Each claim is then grouped into a DRG category, DRG relative weight, APR-DRG category, and APR-DRG severity index within each APR-DRG category and APR-DRG relative weight. For each inpatient claim, a MS-DRG is assigned and a relative weight is determined. The relative weight of DRG, APR-DRG, and MS-DRG indicates the degree of resources used for each case, and is used to adjust the readmission rate across hospitals. Next, index discharges are identified, including all discharged with discharge status of 'alive' from an acute care hospital during the reporting period. Exclusion criteria applied include:

- Discharged deceased
- Discharged to another acute care hospital
- Discharged without a valid patient identifier or hospital identifier
- Discharged with discharge date missing or not valid
- DRG weight or APR-DRG weight missing

Cases with readmission within 30 days from the index discharge date are identified, excluding readmissions related to baby delivery. A logistic regression is run to model a dichotomous readmission probability as a function of case mix of the index discharge cases. The readmission probability for each index case (expected readmission) is calculated, followed by the observed readmission rate for each hospital.[ii] Last, the ratio of expected to observed readmission rate is taken for each hospital. The observed-to-expected (O/E) ratio is indicative of whether more patients or fewer patients were readmitted compared to what are expected based on patient characteristics. If the O/E ratio is less than one, fewer patients were readmitted than expected; whereas if the ratio is greater than one, more patients were readmitted than expected. For example, an O/E ratio of 0.75 would indicate that 25 percent fewer patients were readmitted than expected, whereas a ratio of 1.5 would mean that 50 percent more patients were readmitted than expected.

Risk adjustment methodologies used by State Medicaid Agencies

CMS currently offers a risk adjustment strategy that is traditionally applicable to the Medicare and commercial populations, described in the 2013 Technical Specifications and Resource Manual for the Medicaid Adult Core Set.[19] The methodology uses risk adjustment tables for Medicare and commercial populations developed by NCQA. The risk adjustment strategy proposed by CMS begins by identifying risk adjustment categories for each Index Hospital Stay (IHS), based on discharge condition, the presence of surgeries, comorbidity, age, and gender. Risk adjustment weights are then determined for each IHS based on the same factors, followed by computation of the adjusted probability of a readmission based on the sum of the weights for each IHS.

Currently, five grantee states are applying risk adjustment to the All-Cause Readmission Rate measure. Of the five grantee states where risk adjustment of the measure is underway, methodology differs substantially. Louisiana and Arkansas are using the aforementioned tables published by NCQA geared

---

[ii] Expected readmission rate for each hospital is calculated as the sum of the readmission probabilities for each test case, divided the total number of discharges in the denominator for the hospital

towards commercial populations. Louisiana state representatives cite that NCQA suggested the use of the commercial tables in the interim period before a Medicaid-specific methodology is developed. The state identifies risk adjustment categories based on surgeries, discharge conditions, comorbid conditions, age, and gender. NCQA's published weights for commercial populations are then applied to calculate the adjusted probability of a readmission based on the sum of the weights for each index hospital stay.

Iowa and Montana are executing similar custom strategies developed and conducted by the Lewin Group. First, individuals are assigned to an acuity group for risk adjustment using a retroactive risk score developed using Optum Symmetry Suite's Episode Risk Groups (ERG). ERG assigns a weight on Episode Treatment Groups, which are a group of claims identified as belonging to the same episode of care. These episodes are based on diagnosis codes and National Drug Codes (NDC). These treatment groups are then used to identify the chronic and acute conditions that an individual has, along with the severity of those conditions. The number of conditions, severity, as well as age and gender are then used to calculate the individuals' relative risk compared to the average. To create the acuity categories used in the risk adjustment model, the individual's personal risk score or count of comorbidities was compared to the overall population. If the individual was in the top 5th, 10th, 25th, 50th, percentile or below, they were assigned to the Very High, High, Medium, Moderate and Low acuity groups, respectively.

Observed readmissions were calculated based on age group, sex, and acuity level. The observed readmission rate was then weighted by the percent of total admissions represented by that age, sex, and acuity group. The variance was calculated on the observed readmission rate across acuity groups for each age sex group. The variance is useful in examining whether the O/E ratio represents the age sex group as a whole or if the ratio may have been skewed by one particular acuity group.

An adjusted probability of readmission was calculated by multiplying the probability of readmission for the specific acuity group across all age sex groups by the percentage of admissions in the age, sex, and acuity group. The observed weighted average readmission rate and adjusted probability of readmission were summed by age and sex and then the observed weighted average readmission rate was divided by the adjusted probability of readmission to give an observed to expected ratio for that specific age, and age sex group. The statistical significance of the association between member acuity and readmission was assessed using the p-value. If the p-value for readmission was above 5 percent or a clear increasing or decreasing readmissions trend was not seen with decreasing acuity level bands, the association was considered not statistically significant.

The method used by Iowa and Montana may be particularly useful in identifying specific age and sex groups with higher than expected admission rates after adjusting for risk. This allows the state to focus on the populations that are at highest risk of readmission in their population.

Ohio, the fifth state that opted to risk adjust the all-cause readmission measure, is using the NCQA commercial risk adjustment tables for their managed care populations and the Medicare tables for ABD. Washington, the sixth state, is in the process of developing a Medicaid-specific risk adjustment methodology.

B. Medicare quality measures

Background

CMS has established risk adjustment methodologies for several Medicare quality measures similar to the Plan All-Cause Readmission Rate measure. The Medicare All-Cause Unplanned Readmission Measure for 30 days Post Discharge from Long-Term Care Hospitals and All Cause Unplanned Readmission

Measure for 30 days Post Discharge from Inpatient Rehabilitation Facilities measures the risk-standardized rate of unplanned, all-cause readmission for patients discharged from an inpatient rehabilitation facility (IRF) or long-term care hospital (LTCH), respectively, who were then readmitted to a short-stay acute-care hospital or a long-term care hospital within 30 days of discharge.

Conceptual framework

Both measures track readmission rates for 30 days after a patient is discharged from an LTCH or IRF facility. Readmission rates are a function of patient characteristics and the factors such as communication between providers or between providers and patients; prevention of and response to complications; safeguards for patient safety; and coordination of transition to the next care environment.[29] Elevated readmission rates raise the question of whether the readmissions occurred due to poor care transition processes at discharge or due to the patient's pre-existing comorbidities.

Risk adjustment variables

The complete list of risk adjustment variables is as follows: [29,30]

- Age/sex categories;
- Original reason for entitlement
- Surgery category if present;
- Receiving dialysis in prior short-term stay, defined by presence of revenue code;
- Long-term ventilator patient (for LTCH measure), defined by ICD-9 procedure code;
- IRF case-mix groups (for IRF measure)
- Principle diagnosis on short-term stay bill;
- Comorbidities from secondary diagnoses on the prior short-term bill and diagnoses from earlier short-term stay up to 1 year before admission to LTCH or IRF (clustered using HCC groups);
- Length of stay and length of stay in intensive care in the prior short-term hospital stay (LTCH measure)
- Length of stay in the prior short-term hospital stay (IRF measure); and
- Counts of prior short-term admissions in the 365 days before the LTCH or IRF admission.

Risk adjustment methodology

Generally, the risk-adjusted readmission rate for each type of facility is the mean rate of readmission in the measure population, multiplied by the ratio of the predicted number of readmissions at each facility to the expected number of readmissions for the same patients if treated at the "average" facility.[30] The standardized risk ratio is then multiplied by the mean rate of readmission in the population.

Both measures use a hierarchical regression methodology, within which a logistic regression predicting the probability of a countable readmission is run. The formula is hierarchical because both individual patient characteristics are accounted for as well as the clustering of patients into the respective facility types. The model estimates the average predictive effect of patient characteristics across all facilities of each type as well as effect of each individual facility on the average readmission; facility effects are

assumed to follow a normal distribution. With respect to facility effects, the hierarchical model accounts for known predictors of readmission.

The predicted number of readmission after adjusting for case mix is the sum of the probabilities of readmission in all patients in the facility measure, inclusive of patient and facility characteristics. The expected admission rate for the same patients at an average facility of each type is calculated using the same equation but excluding the facility effects. This risk-standardized predicted-to-expected ratio measures the degree to which readmissions are higher or lower than expected, and may be multiplied by the mean readmission rate to obtain the risk-standardized readmission rate for each individual facility. This procedure is recalibrated for each measurement period to allow the estimated effects of patient characteristics to change over time and in response to medical treatment trends.

C. Minnesota Physician Clinic Quality Measures

Background

A bipartisan health reform law enacted in 2008 called for developing a standardized set of physician clinic quality measures and a system for collecting and reporting data on a subset of those measures for Minnesota. The new law also required formulating a methodology for risk adjusting the measures to use in a quality incentive payment system that accounts for variations in patient population.[31]

In partnership with the Minnesota Department of Health, Minnesota Community Measurement (MNCM) collects data submitted by physician clinics via a web-based portal. The collaborative seeks to minimize administrative burden throughout the data reporting process and aims to collect report quality measure results that are comparable across providers. Clinic-level rates are risk-adjusted for select measures.

Conceptual framework

Optimal Diabetes Care, Optimal Vascular Care, Optimal Asthma Care, and Colorectal Cancer Screening were identified by MDH as measures that may potentially benefit from accounting for patient demographics and other factors that may potentially influence outcomes. Primary payer mix services as a proxy for these factors.

The Depression Remission at 6 Months measure is risk adjusted by severity level amid concern among stakeholders that potential differences in depression severity among patient populations could unfairly affect publically reported results. Concerns were raised that clinics that treat a higher proportion of severely ill patients would have poorer remission rates, under the assumption that severely ill patients are less likely to achieve remission.[31]

Risk adjustment variables

Primary payer mix (i.e., commercial; Medicare; and Minnesota Health Care Programs, uninsured, and self-pay) is the risk adjustment variable for Optimal Diabetes Care, Optimal Vascular Care, Optimal Asthma Care, and Colorectal Cancer Screening measures

Severity level, as reported by initial PHQ-9 score, is the risk adjustment variable for Depression Remission at 6 Months. While primary payer mix was also considered as a risk adjustment variable for this measure, research indicated that primary payer type did not affect the likelihood of proper care once treatment commenced.[31]

Risk adjustment methodologies

Risk adjustment methodologies for reporting for the various types of quality measures are presented below:

Table 3. Physician Clinic Quality Measures—Risk Adjustment by Primary Payer Type.

Primary Payer Type	Quality Measures
• Commercial Insurance • Medicare • Minnesota Health Care Programs/Uninsured/Self-Pay	Optimal Diabetes Care
	Optimal Vascular Care
	Optimal Asthma Care—Ages 5-17
	Optimal Asthma Care—Ages 18-50
	Colorectal Cancer Screening

These three measures are adjusted using a statewide average distribution of patients across three major payer types: commercial; Medicare; and MN health care programs (MHCP) and uninsured/self-pay. Primary payer mix serves as a proxy for variables that reflect differences in patient characteristics and other factors that impact outcomes. According to MNCM, insurance product type can be indicative of a patient's socioeconomic status, and can be used to standardize results.[32] The actual measurement results are multiplied by the statewide distribution of patients by primary payer type. Each clinic's score for each payer type is multiplied by the statewide average distribution by payer to allow clinics with differing payer mix proportions to be comparable to each other for populations within the same payer type.

Issues with low volume of reported data, common among smaller clinics and clinics reporting on a sample basis were addressed in two ways: by combining the MHCP and Uninsured/Self-pay payer categories, and by incorporating the statewide rate into the payer category rate proportional to the number of patients in cases where a physician clinic has fewer than ten patients within a payer category. The first strategy was developed partly in response to the low un-insurance rate in the State and the need to combine payer types to aggregate data; available data indicated that the MHCP enrollees had results most similar to patients in the Uninsured/Self-pay category.

Table 4. Physician Clinic Quality Measures—Risk Adjustment by Severity.

Severity Category	Quality Measure
• **Moderate: Initial PHQ-9 score of 10-14** • **Moderately Severe: Initial PHQ-9 score of 15 to 19** • **Severe: Initial PHQ-9 score of 20 and above**	**Depression Remission at 6 Months**

Rates for the Depression Remission at 6 Months measure were adjusted using a statewide average distribution across three bands of depression severity by initial PHQ-9 score. Backed by research by the

University of Minnesota indicating that Depression remission varies as a function of initial severity and comorbidity, the potential for differences in severity of depression among the patient population was cited for influencing measurement outcomes. For example, clinics treating a higher proportion of severely depressed patients may have lower remission rates because patients with more severe levels of Depression are less likely to achieve remission.

The risk adjusted measurement outcomes assume that all physician clinics have the same distribution of patients across the three severity levels and multiples the actual result for each severity category by the statewide distribution of patients. While primary payer type was also considered as a variable for risk adjustment, research indicated that while payer type may be indicative of access to care, it little correlation to adequacy of care given once treatment has begun.

In the 2013 calculation of the Optimal Diabetes Care, Optimal Vascular Care, Optimal Asthma Care, and Colorectal Cancer Screening measures, approximately ten percent of patient records did not identify the payer type so these records were redistributed for each clinic in the same proportion as their identified counts were originally distributed, in attempt to include all records in the risk adjustment formula.[32] For the Depression Remission at 6 Months measure, a normalizing strategy was applied in cases when a physician clinic has less than ten patients in a severity category. In these cases, the statewide rate is incorporated into the severity category rate, relative to the number of patients in that severity category. The quality rates were then reweighted using the statewide average patient distribution to obtain a final risk-adjusted rate.

D. Risk adjustment in Minnesota Statewide Quality Reporting and Measurement System

Background

Minnesota's 2008 health reform law led to the establishment of the Minnesota Quality Incentive Payment System (QIPS), under which quality-based incentive payments are made to providers based on absolute performance or improvement over time. QIPS has established a uniform statewide pay-for-performance structure with the goal of reducing health providers' burden of accommodating varying types of pay-for-performance programs. The measures selected for quality-based incentive payments includes quality measures for both physician clinics and hospitals that have been identified as priority areas by the community. QIPS uses non-payer-specific market-wide data submitted by physician clinics and hospitals as required by the Minnesota SQRMS.

Pay-for-performance initiatives seek to align payment incentives with provider motivations that drive improvements in health care quality. Risk adjustment is particularly important for these types of programs because a perception that pay-for-performance measures do not effectively account for patient risk factors leaves opportunity of health care providers or facilities to avoid high-risk patients who may adversely impact their performance scores.[6] As a result of this adverse selection, in which providers intentionally avoid medically complex or high risk patients, vulnerable subpopulations of patients may lose access to care.[33]

Conceptual framework

The decision to perform risk adjustment by primary payer type for two quality measures was at least partially dictated by the availability of data. A workgroup led by MNCM and the Minnesota Department of Health concluded that risk adjustment by payer type would be an "acceptable proxy for differences in the severity of illness and socio-demographic characteristics of clinics' patient populations," given data limitations.[32]

The depression remission at six months measure is risk adjusted for patient severity based on research and stakeholder input indicating clinics treating a higher proportion of severely ill patients have poorer remission rates compared to clinics treating less severely ill patients because patients with more severe levels of depression are less likely to achieve remission. Research by the University of Minnesota also indicates that depression remission may vary depending on initial severity and comorbidity. Higher initial severity scores are associated with worse response to treatment.

Risk adjustment variables

Primary payer type (i.e., commercial; Medicare; and Minnesota Health Care Programs, uninsured, and self-pay) is the risk adjustment variable for the Optimal Diabetes Care and Optimal Vascular Care measures.

Severity level, based on initial PHQ-9 severity score, is used to risk adjust the Depression Remission at 6 Months measure.

Risk adjustment methodologies

Table 5. Risk Adjustment Methods for QIPS Quality Measures.

Quality Measure	Risk Adjustment Methods
Optimal Diabetes Care	Primary payer type (i.e., commercial; Medicare; and Minnesota Health Care Programs, uninsured, and self-pay)
Optimal Vascular Care	Primary payer type
Depression Remission at Six months	Patient severity based on initial PHQ-9 severity scores

Theoretically, risk adjusting or population-standardizing quality scores to the average statewide payer mix will account for variations in outcomes that result from differences in patient risk factors that are not under the control of the providers. The Department cites that although more comprehensive approaches are possible, they would require additional data and greater administrative burden for providers. To risk adjust for primary payer type, each clinic's score for each payer type is multiplied by the statewide average distribution of patients in that payer type.

E. Risk adjustment for pediatric quality indicators

Background

Quality of care measurement for children is subject to unique challenges because several factors, such as the need for adult assistance, low morbidity and mortality rates, and specialized pediatric services, complicate the definition of quality measures of pediatric populations.[1]

Conceptual framework

The basic premises of risk adjusting pediatric quality measures is similar that for other subpopulations and the population at large. The goal of risk adjustment is to isolate factors under the control of the provider and differentiate between outcomes that arise due to the provider's actions and those that stem from the patient's inherent characteristics.

Risk adjustment variables

Risk adjustment variables vary depending on the outcome measure and pediatric age group. For example, Clinical Risk Index (CRIB) score uses birth weight, gestational age, the presence of congenital malformations and maximum base access, minimum and maximum appropriate inspired oxygen concentration in the first 12 hours to predict mortality in neonatal units.[34]

Risk adjustment methodologies

Case-mix adjustment method use ICD-9-CM diagnosis codes and/or pharmacy claims from administrative data to predict future costs for a population.[1] Although designed for use in setting capitated payment levels, another application in assessing provider performance.[1] Kuhlthau et al. cites that ACGs, DCGs, CDPS and CRGs have been used for risk adjusting pediatric quality indicators, but no further information about how these methodologies were applied or the risk factors considered could be located. Kuhlthau et al. also points out that these methodologies are particularly beneficial for adjustment of adult quality measures due to their ability to quantify the combined effect of coexisting conditions. By the same token, their applicability to pediatric quality measures is diminished because comorbid conditions do not occur in children as frequently.[35] Case-mix adjustments are also been shown to predict subspecialist visits in both adult and child populations.[36-37] They have been successfully used to predict child hospitalizations and emergency room visits.[38]

Several illness severity measurement systems for risk adjusting neonatal outcomes, such as the Clinical Risk Index for Infants (CRIB), has been used to account for case-mix and adjust infant mortality rates to allow comparison between neonatal units. Crude hospital mortality rates can be adjusted for CRIB using multiple logistic regression to compare performance between units; rankings can then be compared with the indirect standardized mortality ratio.

F. NCQA Health Employer Data and Information Set (HEDIS) & HEDIS Relative Resource Use Measures

Background

NCQA is the national leader in developing and implementing performance and quality indicators for health plans through its Health Employer Data and Information Set (HEDIS) initiative, the most widely used set of performance metrics in the country.[39] HEDIS consists of 80 quality metrics across 5 domains of care, which are voluntarily collected and reported by over 90 percent of health plans to measure and compare performance across plans.[40] Programs such as the Medicare Advantage program and the Federal Employees Health Benefits Program also employ HEDIS. The goal of HEDIS is to provide health plans and consumers with information to compare health plan performance. Measure results are posted in Quality Compass, a web-based tool accessible to the general public.

HEDIS measures cover a wide range of health issues, and focuses on specific, measurable processes of care that have been shown to prevent morbidity in major populations.

HEDIS measure selection and methodology are evaluated annually by NCQA's Committee on Performance Measurement, a group representing health plans, employers, consumers, and other stakeholders. HEDIS applies risk adjustment for certain measures by restricting the denominator to include only those who meet certain criteria for eligibility, instead of including the entire population or condition within a measure. This approach minimizes the need for post-hoc risk adjustment.[36] NCQA also works with health plans to develop adjustment techniques for measures that are most likely to be affected by population risk factors.[41]

Quality measures

In 2010, NCQA introduced the HEDIS Relative Resource Use (RRU) measures in diabetes, cardiovascular disease, hypertension, asthma, chronic obstructive pulmonary disorder (COPD) and low back pain performance, reported through its online Quality Compass: RRU & Quality Index.[41] RRU measures focus on member use of resources, differentiating between variation in resource use and utilization. When applied in tandem with HEDIS quality data, resource use data may help health plans identify further opportunities for improved efficiency and value.[41]

The Relative Resource Use for People with Diabetes measure comprises ten diabetes HEDIS measures:

- Hemoglobin A1c (HbA1c) Testing
- HbA1c Poor Control (>9.0%)
- HbA1c Poor Control (<8.0%)
- HbA1c Poor Control (<7.0%) (only reported for commercial and Medicaid)
- Eye Exam (Retinal) performed
- LDL-C Screening
- LDL-C Control (<100 mg/dL)
- Medical Attention for Nephropathy
- Blood Pressure Control (<130/80 mm Hg)
- Blood Pressure Control (<149/90 mm Hg)

The RRU for People with COPD is reported with the HEDIS COPD quality index:

- Use of Spirometry in the Assessment and Diagnosis of COPD exacerbation; and
- Pharmacotherapy Management of COPD exacerbation; whether a systemic corticosteroid was dispensed within 14 days of the even or whether a bronchodilator was dispensed within 30 days of the event.

The RRU for People with Cardiovascular Conditions is reported with the cardiovascular conditions quality index:

- Cholesterol management for patients with cardiovascular conditions(including LDL-C Screening, LDL-C Control)

- Beta-blocker treatment after a heart attack

Risk adjustment variables

Members' age, gender, and HCC-RRU category (risk marker) are used to determine risk score and risk cohort. HCC-RRU categories are comprised of 75 diseases and conditions.

Risk adjustment methodology RRU measures are risk-adjusted, but do not rely on propriety risk-adjustment tools. Instead, the NCQA risk adjustment model is based on the CMS HCC approach, whereby a member's age, gender, and HCC-RRU category determine their risk score and cohort.[42] Exclusions are made for high-cost clinical conditions (e.g., HIV/AIDS, active cancer, transplantation) and measure-specific co-morbid exclusions.

Health plans receive a report of their RRU results from NCQA with a Plan Population Comparison (O/E Ratio) and a Plan to Plan Comparison (Index Ratio). In the O/E Ratio, observed values indicate actual resources used by the plan, while the denominator, the expected value, is a risk-adjusted benchmark of the resources that the plan was expected to use. The benchmark is based on the expected value in NCQA's estimation of resource use after risk adjustment. Health plans' RRU results are adjusted based on member case mix.[42] For each of the five major clinical conditions, members are assigned to a clinical cohort category. The clinical cohort categories are further stratified, first by the existence or lack of a relevant comorbidity, then by age and gender. Member month and summarized standardized cost are separately reported for each member cohort by health plans. NCQA will then compute the expected per member per month (PMPM) results.

G. NCQA Consumer Assessment of Health Plans Study (CAHPS) Survey

Within HEDIS, The Consumer Assessment of Health Plans Study (CAHPS) survey measures members' satisfaction with health plan accessibility and customer service. CAHPS score results are not currently risk adjusted. However, certain questions about parent and child characteristics are embedded in the survey and could be used to adjust responses. Researchers have found that applying those responses to risk adjust the adult CAHPS data has a small but significant impact on quality scores.[43]

H. Risk adjustment for home health agencies

Background

All Medicare-certified home health agencies collect standardized patient status information through the Outcome and Assessment Information Set (OASIS). OASIS was designed to enable accurate measurement of home health outcomes and adjust for the patient risk factors that affect those outcomes.[19] OASIS provides home health agencies with the Outcome-based Quality Improvement (OBQI) Report, which is derived from OASIS data and measures changes in patent health status over time. The OBQI report includes 37 risk-adjusted outcome measures.

Conceptual framework

Home health agencies use the OBQI measures to track patient health status and monitor the progress of quality improvement initiatives. The comparisons must, therefore, be adjusted for patient risk factor differences (over time for each agency and between the agency and the reference group) because patient outcomes depend on both the quality of care as well as his or her individual risk factors. Risk adjustment

would account for differences in one agency's patients compared to the national reference sample or patients from an earlier time period, thus minimizing the chance that differences in outcomes are misinterpreted as arising from the quality of care provided by the agency.[19]

Risk adjustment variables

272 patient-level attributes are available within OASIS to serve as potential risk adjustment factors.

Risk adjustment methodologies

Multivariate modeling using logistic regression techniques is used. First, the associations between an outcome measure and patient-level risk factors are determined from the 272 patient-level attributes within OASIS. Based on the relationships between risk factors and outcome measure, the predicted outcome values for each patient in the agency is determined; patients' predicted values are then combined to generate an agency-level expected outcome. The expected outcome rate is then compared to the agency's observed rate and compared to the adjusted national reference rate.

Researchers Shaughnessy et al. (2002) conducted a study of risk adjustment of outcome measures for the OBQI reports, whereby separate risk models were created for each of the 41 improvement or stabilization outcome measure in the report, based on data collected from the CMS Outcome and Assessment Information Set (OASIS).[19]

For each outcome, a separate logistic regression model was developed using cases from the developmental sample. Each step of the methodology was developed based on published evidence of successful risk-adjustment techniques. Approximately 150 potential risk factors were identified as candidates for each risk adjustment model. The statistical associations between the individual risk factors and the outcome were examined to identify those factors that were empirically related to the outcome. Following, logistic regression analysis using stepwise variable selection was performed to develop a preliminary risk adjustment model.

Each logistic regression model was then estimated for each outcome of interest. Coefficients and odds ratios for each risk factor were evaluated to determine plausibility. At this juncture of the modeling process, the clinical plausibility and conceptual meaningfulness of the correlation between risk factors and outcome measures and the links between the risk factors as a group were given greater weight than statistical considerations. Factors that did not meet these criteria or those with "clinically questionable" coefficients were removed or replaced by other risk factors. In an iterative process, each model was re-assessed several times until a clinically- and statistically-sound model was achieved. The explanatory power of each model was then tested using the validation sample by selecting 20 random subsamples from the sample, calculating a predicted outcome value for each unit and comparing the predicated value to the observed outcome value. This generated one R^2 statistic for each random validation subsample, a measure of the proportion of the outcome variance that can be explained by the model. C-statistics were included alongside the R^2 statistic in later iterations to identify cases with high c-statistic despite a low R^2 value, which is can also signify a potentially useful model. Instances of a large difference between the developmental sample R^2 and the average of the 20 subsample R^2s, or in cases of an exceptionally large range in the subsample R^2s indicate that the model had been "over-fitted" to the developmental sample. This required re-estimation of the model, re-evaluation of risk factors' suitability, and re-calculation of model coefficients in an iterative process until a stable model is established.

I. Risk adjustment of CMS quality measures for nursing homes

Background

The Centers for Medicare and Medicaid Services publishes Nursing Home Compare, a quality report card providing information on about 19 clinical nursing home quality measures, geared towards patients, families, policy makers, and with the goal of informing quality improvement efforts by nursing homes.

Conceptual framework

As in most other health care environments, health outcomes in nursing homes are a function of not only quality of care, but also patients' risk factors and inherent characteristics. Raw incidence or prevalence rates of adverse outcomes therefore cannot be used as standalone units of comparison between nursing homes.

Previous research has also indicated that reported performance using outlier designations and nursing home facility rankings based on raw outcomes differed as the degree of the risk adjustment was applied. This risk adjustment test was first performed based only on age, then increased to age and a case-mix index adjustment, then further increased to a more comprehensive risk adjustment methodology that included several factors on patients' physical and mental health.[44] Other studies on quality measurement in veterans' nursing home and long-term care facilities have yielded similar results.[45-46]

Risk adjustment variables

For 14 of the 19 CMS Quality Measures, risk adjustment is performed through application of one or several exclusion criteria.[iii] For the remaining five measures, covariate risk adjustment is performed in addition to applying exclusion criteria.

Variables used in the covariate risk adjustment include:

[iii] Exclusion criteria include:
(1) target assessment is an admission;
(2) the QM did not trigger (resident not in numerator) and/or there is missing or inconsistent data on Minimum Data Set (MDS) items required for quality measure
(3) MDS item cannot show decline because it has total dependence value if activity did not occur;
(4) resident is comatose or comatose status is unknown;
(5) resident has end-stage disease or receives hospice care;
(6) not qualify as high-risk and bed mobility or transfer or comatose is unknown;
(7) pain symptoms are unknown or inconsistent on 14-day assessment;
(8) pressure ulcers are missing in the 5-day / 14-day assessment;
(9) not in facility during influenza season or facility unable to obtain vaccine;
(10) not eligible for given vaccine or resident offered vaccine but declined;
(11) the Mood Scale score is missing on the target or prior assessment;
(12) the Mood Scale score is at a maximum (value 8) on the prior assessment;
(13) high risk residents with severe cognitive impairment or totally dependent in mobility ADLs;
(14) resident has an indwelling catheter or an ostomy;
(15) not qualify as high risk and the cognitive impairment items or any of the mobility ADLs unknown;
(16) locomotion on unit is unknown or shows some dependence on target or prior assessment;
(17) urinary tract infection in last 30 days is missing on target assessment;
(18) weight loss is missing or resident is receiving hospice care or hospice status unknown;

Table 6. Variables Used in Covariate Risk Adjustment.

CMS Quality Measure	Covariate Risk Adjustment Variable(s)
Percent of Long-Stay Residents Who Have Moderate to Severe Pain	Independence or modified decision making ability on prior assessment variable
Percent of Long-Stay Residents Who Have/had a Catheter Inserted and Left in Their Bladder	Bowel incontinence, pressure ulcers on the prior assessment
Percent of Long-Stay Residents Whose ability to move about in and around their room got worse	Recent falls, extensive support or more dependence in eating / toileting on the prior assessment
Percent of Long-Stay Residents With delirium	Prior residential history
Percent of Long-Stay Residents With pressure ulcers	History of resolved pressure ulcers, limited assistance in bed mobility, bowel incontinence, diabetes or peripheral vascular disease, and Low Body Mass Index

Risk adjustment methodologies

Current CMS methodology to calculate most nursing home quality measures uses limited risk adjustment, applied mostly through exclusion criteria, intended to create more homogenous comparison groups. CMS' decision to apply minimal risk adjustment came after a series of studies that examined quality measured applied with varying degrees of risk adjustment, ranging from those based mostly on exclusion criteria to measures that have undergone extensive risk adjustment based on multivariate regressions. Other methodologies have also been explored.

The methodology used by Mukamel et al., 2008 in assessing the impact of more extensive risk adjustment on the on the quality measures published by CMS found that there are additional risk factors beyond those currently included in CMS methodology that are significant to the health outcomes of interest.

The researchers first refined the outcome measure by applying the same inclusion and exclusion criteria as the CMS methodology. All outcomes were defined as dichotomous variables, set to 1 if patient experienced the outcome, 0 if otherwise. Next, additional risk factors beyond those already included in the CMS methodology were identified from the literature. For each outcome, risk factors that considered with accessible data were evaluated by a geriatrician familiar with the nursing home population to assess their likelihood of influencing the outcome measure of interest.

Mukamel et al. estimated a hierarchical random effect logistic model to identify each risk factor's association with the relevant outcome. The models allowed for potentially non-random clustering of residents within facilities. The possibility of an efficiency advantage gained by pooling data was not relevant due to the very large sample size. The model was guided by inspection of the c-statistic, which assesses the discriminatory power of the model, and the Hosmer-Lemeshow statistic, which measures the calibration of the model. As this juncture, the researchers discovered that the model fit is improved when

the risk models are stratified by age, so separate models were estimated for each of four age groups. This provides the same effect as interacting age with all other risk factors.

Multiple random samples of 100,000 observations were created due to the very large sample size. The average of each estimated coefficient across 200 iterations was calculated for all samples. The change in the average coefficients between 150 to 200 iterations was less than 2 percent for all samples, which suggested that the coefficient estimates were converging to their true means within a small margin of error. The average coefficients were used to calculate the expected probability of the outcome for each nursing home resident, conditional on his or her specific risk factors. The expected and observed facility rates were found by taking the average of the individual expected probability and the average of the observed outcomes, respectively. The final outcome was then calculated as the ratio of the observed to expected rate.

J. Hospital and surgery quality measures: Coronary Artery Bypass Graft Mortality Rate

Background

One of the areas of care most developed with regard to risk adjustment and provider profiling in quality assessment for coronary artery bypass grafting (CABG), the standard surgical treatment for patients with coronary artery disease who do not responds to medical treatment and/or for whom angioplasty has proven ineffective. The CABG rate measures the number of coronary artery bypass graft surgeries performed per 1,000 members 45-64 and 65+ years of age. A number of states initiated reporting of CABG performance data in the 1990s to assist patients and providers in selecting hospitals and surgeons for the procedure. Since then, there has been conflicting evidence regarding the effects of public reporting, with some sources claiming that reporting is associated with quality improvements at the hospital and physician levels, while others studies claim that such reporting lead to preferential selection of healthier patients and avoidance of high-risk patients and those with higher comorbidities.[47,48]

Conceptual framework

A uniformly high CABG rate may indicate that a health plan or hospital either services a population with an unusually high rate of severe heart disease or that other medical approaches to treating coronary heart disease are underutilized or poorly utilized.[41] High CABG rates coupled with low angioplasty rates suggest invasive procedures may be applied prematurely in the treatment process. Conversely, an unusually low rate of CABG may raise questions about the availability of such services. Risk adjustment of the measure is necessary to differentiate between the implications of these outcomes. Risk adjustment for this measure is performed in a variety of ways.

Risk adjustment variables

Patient demographics, clinical characteristics, and all other measurable, prognostically important patient risk factors present prior to CABG surgery. Hospital characteristics (i.e., case volume, nurse-patient ratio, etc.), procedures prior to surgery (i.e., intra-aortic balloon pump, carotid screening) or after surgery (therapy, discharge planning, etc.) are not adjusted for because these factors are under the control of the provider.[41, 49]

Risk adjustment methodologies California CABG Outcomes Reporting Program

The California CABG Outcomes Reporting Program (CCORP) uses isolated CABG (defined as CABG in which no other major cardiac surgery is performed) and risk-adjusted mortality rates (RAMR) to provide annual public reports for all nonfederal hospitals and surgeons who perform the procedure. The state's goal for risk adjusting the CABG measure is to adjust the observed outcome rates so that they reflect the average illness severity of a hospital's patient mix.[49]

The current statewide methodology entails generating an expected mortality for each patient. This is done by first entering CCORP registry data into a logistic regression model, which predicts operative mortality and evaluates every risk factor's influence on mortality. Expected mortality is used as a proxy for case mix, probability of death, and average illness severity. Next, a ratio of each hospital's observed to expected (O/E) mortality rate is calculated, which is then multiplied by the state's average observed mortality rate to arrive at the hospital's risk-adjusted mortality rate.

An earlier study was performed to assess the impact of public reporting of isolated CABG outcomes in California by comparing CABG volume, patient operative risk, and operative mortality for hospitals and surgeons in 2003, the first year of the state-mandated public reporting, to outcomes in 2006.[50]

The researchers initially included five demographic variables, nine comorbidities, four hemodynamic variables, and seven cardiac or prior intervention variables to compare difference in patient preoperative profiles between 2003 and 2006. A multivariable logistic regression to calculate predicted mortality as an index for patient preoperative case mix was used to compute a risk-adjusted mortality rate for each hospital and surgeon. Combined 2003 and 2004 CCORP data was used to avoid bias. Patients were assigned to quintiles based on their predicated operative mortality, and a general linear model was used to compare variance between quintiles. The researchers then computed the provider-specific risk-adjusted mortality rate (RAMR) based on hospital and surgeon observed and predicted operative mortality. These steps were performed for 2003 and 2006 data to enable comparison of volume, provider-specific predicted and risk-adjusted mortality between the two years.

In other reports assessing the quality of medical providers for surgical procedures and medical conditions that are released by federal and state government, quality assessment has been primarily performed by adjusting providers' mortality rates on the basis of severity of illness of patients prior to treatment.[47] The methodologies used to risk-adjust mortality in these reports were largely centered on standard logistic regression, which models the occurrence of a binary outcome on the basis of patient risk factors. The expected mortality rate for each provider is obtained by summing the individual predicted probabilities of death for all of the provider's patients and dividing by the total number of patients. This is then compared to the observed mortality rate to calculate a RAMR for each individual provider.

Hospitals participating in the CCORP raised several concerns regarding the methodology used for risk adjustment.[51] One, in particular, was regarding the variation in coding of the risk factors, which can easily affect the validity of the risk-adjusted results. The hospital raised the point that overstating patient risk profiles can provide an unfair advantage to some hospitals, while not including certain difficult-to-measure risk factors in the model was sure to impact a patient's risk of exhibiting the outcome.

New York's experimentation with hierarchical statistical modeling

In the past decade, some researchers have advocated the use of hierarchical statistical models, also known as multilevel models or random-effects models, to analyze nested data. They point to the standard logistic

regression's inability to account for the statistical impact of the clustering of patients within hospitals, potentially leading to inaccurate estimates due to the small sample sizes. The hierarchical model counters these effects by weighting the provider performance estimates towards the overall mean, with more weight applied to lower volume providers.

As of 2005, the New York State CABG surgery reports used logistic models to assess current provider performance. However, due to the two-year time lag between reports, some argued that the report results might also be interpreted as a prediction of future performance. Hannan et al. (2005) performed a comparison of the results of a standard logistic regression to those of a hierarchical logistic model, from the perspective of predicting risk-adjusted mortality rates for CABG surgery in the state. The development of the logistic regression model followed standard procedure, known as "indirect adjustment," with each patient's predicted probability of death calculated using stepwise logistic regression to identify patient risk factors that are most closely associated with mortality and should therefore be weighted more heavily. Comparison of the expected mortality rate to the observed mortality rate was performed by taking the ratio of the observed to expected rates and rescaling by the multiplying by the statewide mortality rate to obtain the RAMR. Risk-adjusted rates that were significantly higher or lower than the statewide rate were identified as high or low outliers, respectively.

The hierarchical logistic model employed a random intercept model without second-level covariates. Individual patients represented the first level, while hospitals formed the second level. The first level covariates were the same ones used for the standard logistic model in that given year. This methodology was developed in reference to previous approaches used for risk-adjusted mortality rates.[52-53] Unlike in the logistic regression model, the dependent variable in the hierarchical model is dependent on the individual hospital. The researchers note that to maximize the predictive power of the hierarchical methods, they should be as "exchangeable" as possible, based on the assumption that different hospitals will have similar risk-adjusted outcomes over time. In contrast, hospital-level characteristics that may be related to long-term outcomes should be included as second-level covariates instead.

The results of the study suggest that the performances of the models in predicting mortality rates were similar; no significant differences were identified. Differences in hospital mortality rates accounts for a minimal portion of the unexplained variation in mortality.

K. Hospital and surgery quality measures: AHRQ Quality Indicators

Background

AHRQ Inpatient Quality Indicators (IQIs) measure quality of care in hospitals using inpatient administrative data. The set includes measures for inpatient mortality rates for medical conditions, inpatient mortality rates for surgical procedures, hospital-level procedure utilization rates, area-level utilization rates, and procedure volume measures.

Conceptual framework

In response to increasing interest among hospitals for ways to evaluate and evaluate performance with respect to a group standard, AHRQ established the IQIs to allow hospitals to compare their performance in several areas (mortality, utilization, and volume) to expected outcomes in the AHRQ Healthcare Cost and Utilization Project (HCUP) reference databases. Risk adjustment is needed to standardize performance outcomes across hospitals.

Quality measures

Morality rate measures for conditions: acute myocardial infarction (AMI), AMI without transfer, congestive heart failure, gastrointestinal hemorrhage, hip fracture, pneumonia, acute stroke[54]

Mortality rate measures for procedures: abdominal aortic aneurysm repair, coronary artery bypass graft, craniotomy, esophageal reaction, hip replacement, pancreatic resection, percutaneous transluminal coronary angioplasty, carotid endarterectomy

Hospital-level procedure utilization rate measures: cesarean section delivery, primary cesarean delivery, vaginal birth after cesarean (uncomplicated), vaginal birth after cesarean (all), incidental appendectomy in the elderly, bilateral cardiac catheterization, laparoscopic cholecystectomy

Area-level utilization rate measures: coronary artery bypass graft, hysterectomy, laminectomy or spinal fusion, percutaneous transluminal coronary angioplasty

Volume of procedures measures: abdominal aortic aneurysm repair, carotid endarterectomy, coronary artery bypass graft, esophageal resection, percutaneous transluminal coronary angioplasty

Risk adjustment variables

Risk adjustment performed on the basis of severity of illness or risk or mortality, after adjusting for age and gender.

Risk adjustment methodologies

The statistical model used provides parameter estimates for each quality indicator adjusted for age and gender by using APR-DRGs, a classification methodology developed by 3M that allows for the risk adjustment of quality indicators on severity of illness or risk of mortality. AHRQ uses a simple logistic regression modeling approach that assumes that all patient responses are independent and normally distributed. Risk-adjusted rates estimate performance under the hypothetical scenario that each provider had a case mix that mirrored the national average, given their actual performance. The risk-adjusted rate is calculated by taking the ratio of the observed to expected rates at the provider level and adjusting by the observed national average rate.

Alternative statistical models were investigated by AHRQ, including adjusting for within-provider correlation to account for possibility that responses of patients within the same hospital may be correlated even after adjusting for severity of illness, risk of mortality, and basic demographic characteristics. The additional factors that may influence patient results include hospitals' unique medical culture comprised of a combination of factors relating to staff composition and policies. Fitting a simple logistic model to the data that is correlated may result in biased standard errors of the parameter estimates.[55] AHRQ also researched adjusting the models for the anticipated positive correlation among patient responses from within the same facility through use of Generalized Estimating Equations (GEE) approaches and Generalized Linear Mixed Modeling (GLMMIX) approaches, both available through SAS. However, both models are subject to convergence issues and model misspecification that is common among iterative modeling methods.

L. Risk adjustment of other hospital quality measures and surgery quality measures

Background

Research has demonstrated that different risk adjustment methods lead to different quality rankings of hospitals. In a study to determine whether decreasing the number of variables used in the American College of Surgeon's National Surgical Quality Improvement Program (ACS NSQIP)'s procedure-specific outcome measurement, Dimick et al., compared the ability of a full risk adjustment model (containing 21 variables), an intermediate model (12 variables), and a limited model (5 variables) to predict patient outcomes and to risk-adjust hospital outcomes.[56]

Conceptual framework

The study was performed to assess the performance of parsimonious risk adjustment models in comparison to models that adjust for a greater number of patient characteristics.

Quality measures

The researchers used data from the 2005-2007 American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP), which is a clinical registry intended to feedback risk-adjusted performance outcomes, primarily on mortality, to hospitals to assist in quality improvement initiatives.

Risk adjustment variables

The ACS-NSQIP collects over 130 patient and operative variables, including patient demographics, preoperative risk factors, patient laboratory values, intraoperative variables, and postoperative 30-day morbidity and mortality. The researchers chose among these variables to develop risk adjustment models.

Risk adjustment methodologies

The full risk-adjustment model was built by entering all patient-level variables into a stepwise regression model, and preliminarily retaining all variables with $P<0.1$. Further analyses were performed for mortality and morbidity using logistic regression. For the intermediate model, stepwise logistic regression models were ran with all potential risk-adjustment variables included ($P<0.1$ for entry and exit into the model), however, the output of these models included a rank of order of importance. Order of importance was proxy for order of entry into the model, which reflected how strongly the factor was correlated to the outcome variable. For the most parsimonious model, the five earliest entry variables into the intermediate stepwise regression model were combined.

To assess patient-level risk-adjustment, the researchers evaluated the c-index for all three models. The c-index is the area under the Receiver Operating Characteristic (ROC) curve, which reflects the ability of the model to discern patient outcome, ranging from 0.5 (no discriminating ability) to 1.0 (perfect discriminating ability). Calibration was assessed using the Hosmer-Lemeshow statistic, which compares observed and predicted outcomes across the risk spectrum. The statistic only assesses whether the average and predicted rates are similar within subgroups of cases, and does not check for whether the average of the predicted outcomes approximately equals the average of the actual outcomes. The Spearman correlation coefficient to estimate patients' predicated probability of mortality was also assessed for each model.

For this study, the intermediate and limited risk-adjustment models demonstrated similar predicative ability as the full model. The findings were consistent with earlier research concluding that risk adjustment models could be simplified to streamline the data collection process.[57] This research compared a 6 variable risk-adjustment model to a 12 variable model and found that the c-index increased only minimally with the inclusion of the six additional variables.

Summary of Risk Adjustment Methodologies from the Environmental Scan.

Category of measure	Quality measure(s)	Risk adjustment methodology or techniques
Initial Core Set of Health Care Quality Measures for Adults	Plan All-Cause Readmission Rate Measure	▪ Louisiana and Arkansas are using risk adjustment tables for commercial populations developed by NCQA ▪ Ohio is using NCQA commercial risk adjustment tables for their managed care population and Medicare tables for the ABD population ▪ Iowa and Montana are using a custom Medicaid-specific risk adjuster that uses Episode Risk Groups (ERGs) to define acuity groups ▪ Washington is developing a custom Medicaid-specific risk adjustment methodology
Medicare quality measures	All-Cause Unplanned Readmission Measure for 30 Days Post Discharge from Long-Term Care Hospitals; All-Cause Unplanned Readmission Measure for 30 Days Post Discharge from Inpatient Rehabilitation Facilities	▪ Both measures use hierarchical regression methodology (accounts for both individual patient characteristics as well as clustering of patients into the respective facility types)
MN Physician Clinic Quality Measures	Optimal Diabetes Care; Optimal Vascular Care; Optimal Asthma Care	▪ Risk adjustment by primary payer type (commercial; Medicare; and Minnesota Health Care Programs, uninsured, and self-pay) to proxy for variables that reflect differences in patient characteristics and other factors that impact outcomes
MN Physician Clinic Quality Measures	Depression Remission at 6 months	▪ Risk adjustment by severity category, as determined by initial PHQ-9 score

Category of measure	Quality measure(s)	Risk adjustment methodology or techniques
Pediatric quality measures	Neonatal Outcomes; Pediatric ICU Outcomes; Surgical Mortality for Congenital Health Disease	▪ Adjusted Clinical Groups (ACGs) ▪ Diagnostic Cost Groups (DCGs) ▪ Chronic Illness and Disability Payment System (CDPS) ▪ Clinical Risk Groups (CRGs)
HEDIS measures	Unclear which HEDIS measures are risk-adjusted	▪ Applying exclusion criteria for eligibility ▪ NCQA works with health plans to develop risk adjustment techniques
HEDIS RRU measures	RRU measures in diabetes, cardiovascular disease, hypertension, asthma, COPD, and low back pain performance	▪ NCQA-developed risk adjustment model based on the CMS Hierarchical Condition Category (HCC) approach, in which a member's age, gender, and HCC-RRU category determines his/her risk score and cohort
CAHPS measures	Measures are not currently risk adjusted, but survey questions about parent and child characteristics could be used to adjust responses	▪ Researchers have found that applying the parent and child characteristics to risk adjust the adult CAHPS data has a small but significant impact on quality scores
Home health agency quality measures	41 improvement or stabilization outcome measures in the home health agency outcome-based quality improvement (OBQI) reports	▪ Logistic regression, based on an initial set of a 150 potential risk factors
Nursing home quality measures	19 clinical nursing home quality measures in CMS' Nursing Home Compare report card	▪ Limited risk adjustment, applied mostly through exclusion criteria ▪ Five of 19 measures are further adjusted using limited covariate risk adjustment ▪ Researchers used a hierarchical random effect model to investigate the effects of including additional risk factors

Category of measure	Quality measure(s)	Risk adjustment methodology or techniques
Hospital and surgical quality measures	Coronary Artery Bypass Graft Mortality Rate	▪ Hospitals participating in the California Coronary Outcomes Reporting Program (CCORP) primarily use standard logistic regression models. ▪ Research study assessing the impact of the CABG reporting used a multivariable logistic regression with five demographic variables, nine comorbidities, four hemodynamic variables and seven cardio or prior intervention variables ▪ New York CABG surgery reports used logistic regression models ▪ Research study (Hannan et al., 2005) compared a standard logistic regression model to a hierarchical logistical model for NY CABG outcomes, but found results to be similar
Hospital and surgical quality measures	AHRQ Inpatient Quality Indicators (IQIs)	▪ APR-DRGs ▪ AHRQ also investigated using Generalized Estimating Equations (GEE) and Generalized Linear Mixed Modeling (GLMMIX) approaches

Choosing a Risk Adjustment Methodology

A. Selection of risk factors

Thorough identification and evaluation of patient factors that could potentially impact the targeted outcome or process of care is critical to the success of any risk adjustment method. Potential patient risk factors, grouped by dimension, are presented in Table 7. Smith et al. (2009) suggests that a conceptual model with a complete set of risk factors should be initially developed independently of practical concerns, such as the availability of data or the feasibility of data collection. Although it may be impractical to adjust for all pertinent characteristics, it is nonetheless important to identify all potentially significant factors. The capability of attributing residual differences in performance to their root causes, whether through formal risk adjustment or caveating findings, is critical to accurately interpreting performance comparisons across providers or facilities.[2]

Table 7. Potential Patient Risk Factors.

Category and Risk Factor	Available in DHS Data Warehouse[iv]
Demographic characteristics	
Age	X
Gender	X
Race/ethnicity	X
Clinical factors	
Disability status	X
▪ Physical functioning	
▪ Vision, hearing, speech functioning	
▪ Cognitive functioning	
Acute physiological stability	
Principle diagnosis	
Severity of principle diagnosis	
Extent and severity of co-morbidities	
Mental illness, emotional health	
Health-related behaviors and activities	
Tobacco use	
Alcohol and illicit drug use	
Diet and nutrition	

[iv] Sociodemographic data collected for calendar year 2011 by the DHS Data Warehouse include: age, gender, race, disability status, interpreter needed, primary language, income, educational level, county of financial responsibility, county of residence, living arrangement, household size, primary health program, and number of months enrolled.

Category and Risk Factor	Available in DHS Data Warehouse[iv]
Physical activity and exercise	
Obesity and overweight	
Socio-economic/psychosocial factors	
Educational attainment	X
Language(s) spoken	X
Economic resources	X
Employment and occupation	
Familial characteristics and household composition	X
Housing and neighborhood characteristics	X
Health insurance coverage	X
Cultural beliefs and behaviors	
Religious beliefs and behaviors	
Attitudes and perceptions	
Overall health status and quality of life	
Preferences, values, and expectations for health-care services	

Sources: Smith PC, Mossialos E, Papanicolas I, Leatherman S, 2009. Performance measurement for health system improvement. Cambridge University; Risk adjustment for measuring health care outcomes. Third edition. Chicago, IL: Health Administration Press.

B. Risk adjustment and socioeconomic risk factors

The 2012 Health Care Disparities Report for MHCP reported health care performance outcomes for patients enrolled in the managed care components of the state's Medical Assistance and MinnesotaCare programs. MHCP enrollees tend to be of lower socioeconomic status and represent a disproportionate number of persons of color, American Indians, persons with disabilities, and elders.[58] Furthermore, it is recognized that these patients may experience other, often undocumented, hardships that prevent them receiving the appropriate care and services.

The report findings highlighted the differences in outcomes between patients enrolled in MHCP and patients enrolled with managed care programs of commercial payers, employer-based health care insurance, or Medicare programs at medical group and statewide levels. For six of the thirteen measures profiled, outcomes are also compared at the clinic level. The results indicate that statewide gaps in performance outcomes between MHCP and other purchasers narrowed over time for five measures, but have widened over time for six measures.[v] Additionally, differences exist between racial groups within

[v] Statewide gaps in performance rates between MHCP and other purchasers have narrowed over time for the following measures: Optimal Asthma Care—Children Ages 5-17, Appropriate Testing for Children with Pharyngitis, Breast Cancer Screening, Cervical Cancer Screening, and Childhood Immunization Status. Statewide gaps have widened over time for the following measures: Optimal Diabetes Care, Controlling High Blood Pressure, Optimal Asthma Care—Adults Ages 18-50, Appropriate Treatment for Children with URI, Colorectal Cancer Screening, and Chlamydia Screening.

the MHCP patient population for seven HEDIS measures.[vi] As of February 2014, The National Quality Forum (NQF) is examining the issue of sociodemographic factors in the context of risk adjustment of outcome performance measures. While outcome measures are determined in part by the quality and effectiveness of services and treatments, they may also be influenced by patient health status and sociodemographic factors. The guiding principle of risk adjusting for sociodemographic factors is that measurement influenced by factors unrelated to care received needs to be adjusted for relevant case mix differences to avoid inaccurate or misrepresentative inferences about actual performance.[7] The NQF Expert Panel has recently put forth a draft of ten recommendations on risk adjustment using socioeconomic variables.

NQF staff and the Expert Panel have proposed many guidelines aimed to standardize the risk adjustment process. Particularly, NQF recommends that the same considerations for selecting clinical and health status risk factors for risk adjustment be applied to sociodemographic factors. Factors should be not confounded with quality care, should not be highly correlated or redundant with other risk factors, and should be present at the start of treatment. The selected sociodemographic variables should also have face validity and acceptability. NQF lists income, education, homelessness, English language proficiency, and insurance status as sociodemographic factors that should be tested as potential risk adjustment factors.

NQF notes that income information can be difficult to collect in private clinic settings, is not easily encompassed by one question, and can be interpreted differently by geographic region. Further, patients many not feel at ease in supplying income information. As a proxy for income, Federal poverty level exhibits certain advantages, such as a standard definition and widespread application of the measure. Household income may be more meaningful than individual income, but would require simultaneously capturing information on household size. Alternatively, Medicaid status and SSI information is relatively easy to collect, but eligibility status definitions are not consistent across states. NQF does not recommend using "race/ethnicity" as a proxy for SES due to the potential for bias.

Additionally, NQF states that risk adjustment for sociodemographic factors should preserve the ability to identify and address disparities. This may be achieved by reporting risk adjusted results in the context of peer groups based created based on the key sociodemographic variables, and reporting metrics with and without risk adjustment.

Sociodemographic characteristics are widely used in risk adjustment of HEDIS measures. HEDIS adopted the low birth weight (LBW) measure as a quality indicator for effectiveness of care and it is one of the few outcome indicators in the HEDIS set. Low birth weight is an indicator of maternal and child health because it is associated with adverse infant outcomes.[59] In a study to evaluate whether adjusting the HEDIS low birth weight (LBW) measure for material risk factors improves its validity as a quality indicator, Inkelas et al., 2000 identified potential risk adjustment factors that may be associated with the LBW indicator, and used vital records data to develop and test several risk adjustment strategies.[59] The researchers used logistic regression to determine the impact of the mother's risk factors on the probability of having a LBW infant.

The criteria used to select potential risk adjusters found in the literature include: (1) the association with LBW is well-established in the literature (2) the magnitude of the effect found for the factor (3) whether the effect of the risk factor could be diminished by an increase in services provided in the prenatal period. Therefore, ideal risk adjusters are factors that are empirically and abundantly established, demonstrated a significant causal effect of the probability of LBW, and that could not be modified by the health care

---

[vi] Controlling High Blood Pressure, Appropriate Treatment for Children with Upper Respiratory Infection, Appropriate Testing for Children with Pharyngitis, Breast Cancer Screening, Cervical Cancer Screening, Chlamydia Screening in Women, and Childhood Immunization Status (Combo 3).

provider. The researchers also note that other potential risk adjustment factors are also not modifiable and are strongly linked to LBW, but may be controversial. These factors are generally sociodemographic in nature, such as ethnic group and income. Inkelas and colleagues cites that risk adjustment for these factors could mask inherent inequities in service delivery, while not adjusting for the factors could jeopardize the equity of service provided by the incentivizing providers to enroll and provide services to lower-risk mothers. When sociodemographic factors were modeled, it was found that combining obstetric and socioeconomic status (SES) risk factors did not change the odds ratio. While the socioeconomic indicators affected the probability of the LBW outcome, they did not change the relationship between the obstetric variables and the LBW outcome. Other results from the simulation indicated that the risk adjustment effect was small for the most parsimonious risk adjustment model. The mean adjustment using only obstetric risk factors produced a mean change of 0.17 percentage points for the hospital groups included in the study. When socioeconomic status factors and/or maternal race were included in addition to the obstetric factors, the mean adjustment and range increased.

C. Data sources used for risk adjustment

Administrative data, clinical data, or information directly from patients and consumers are primary sources of data used in risk adjustment. Though the conceptual ideal is to have complete information on all patient risk factors, some compromise is generally necessary.

Administrative data

Administrative data is easily obtained at relatively low cost due to its ubiquitous use for administrative purposes such as claims, billing, and record-keeping. This type of data can cover large populations in a uniform manner and is able to define and record using consistent rules and standards. This enables researchers to link patient records longitudinally over time.

Despite widespread use, risk adjustment methods that use administrative data are subject to several possible pitfalls. For example, if the same administrative records are used for reimbursement for payment-related incentives, the data could potentially be skewed to benefit providers. Inaccurate hospital coding of diagnoses may also be present. In a study of 991 discectomy cases admitted to California hospitals, Romano et al., 2002 found hospital codes only exhibit 35 percent sensitivity for the identification of any complications of care, with underreporting markedly worse at hospitals reported to have lower risk adjusted rates.[60] Further, although ICD codes are intended to capture the full range of health conditions and disease, they do not include certain clinical parameters linked to disease severity, nor are they telling of functional status.[61]

Medical records or clinical data

At times, coded administrative data does not provide sufficient clinical content or clinical validity. ICD codes do not sufficiently capture clinical risk factors in certain non-acute care settings, such as nursing home or home health agencies. Medical records or electronic systems containing clinical information can be repositories of such needed clinical data for risk adjustment. However, these data can be expensive to use and may be limited, particularly in outpatient settings, where medical records tend to be highly variable in accuracy in completeness.[2] Further, clinical data is not immune to manipulation or breaches of data integrity. For example, certain functional assessments are made in a clinicians' office or at the bedside by examination of the patient. Because these records are fairly subjective, it may be possible for clinicians to exaggerate a patient's functional impairment when faced with incentive to do so.

Direct information from patients

When certain quality measures seek to gauge patients' perceptions, satisfaction with care, or self-reported functional status, information directly from patients can be a valuable resource for risk adjustment. This information can be obtained by collecting survey data. For example, research indicates that patients with poorer health systematically report lower satisfaction with their health care and providers than healthier individuals.[62] Overall health status is therefore an important risk factor to take into account when interpreting satisfaction surveys. When using direct information, however, it is necessary to consider that factors such as misunderstandings, inaccurate memories, cognitive impairments, and language difficulties may compromise the accuracy of self-reported patient information. Possible respondent bias due to an overall declining response rate may also need to be considered.

D. Performance of the most widely used risk adjustment systems

Table 8 presents a comparison of the required data elements and performance metrics of the leading commercially-available risk adjustment systems on the market. Please note that measures of predictive power are based on cost, not quality. Current, comparisons of predictive power for the models based quality are not yet available.

Table 8. Data Elements and Performance Assessment of Select Risk Adjustment Models.

Risk adjustment Models	Demographic Information	Diagnosis Codes	NDC Codes	Procedure -Revenue Codes	Provider Type	Model Performance (R2) as offered	MAPE (%)
Chronic disability payment system(CDPS)	X	X				12.4	95.8
Chronic disability payment system (CDPS) - Rx	X	X	X			22.0	N/A
Medicaid Rx	X		X			12.9	90.2
Clinical Risk Groups (CRG)	X	X	X	X		14.9	91.4
Adjusted Clinical Groups (ACG)	X	X	X	X	X	16.2	90.4
DCG	X	X				17.4	88
DCG-RX groups	X		X			16.8	85.9

Risk adjustment Models	Demographic Information	Diagnosis Codes	NDC Codes	Procedure -Revenue Codes	Provider Type	Model Performance (R2) as offered	MAPE (%)
Episode Risk Groups (ERG)	X	X	X	X		16.2	87
Pharmacy Risk Groups (PRG)	X		X			17.2	86.4
Impact Pro	X	X	X	X	X	21.3	82.4

The Society of Actuaries periodically publishes reports comparing the performance of selected models. In the most recent report, released in 2007, the models were compared using two standard measures of performance: the mean absolute prediction error (MAPE), and the R^2 measure.[21] MAPE is defined as the ratio of the absolute value of the prediction error to the sample size, where prediction error is the difference between actual medical costs and predicted costs. Like R^2, the MAPE is a single summary measure of predictive accuracy, but it has the advantage of being less sensitive to large claims because it does not the square the prediction error. Often, as in the SOA study, MAPE is expressed as a percentage of the average PMPY cost to create a standardized scale between the models of interest. The formula used is:

$$MAPE = (\Sigma|\text{Actual Costs-Predicted Costs}|) / (\text{Sample Size})$$

This particular evaluation for cost found that the MEDai methodology produced the highest R^2 and lowest MAPE among all models in the study (including all featured methodologies in Table 8 and others), while the DCG model produced the highest R^2 lowest MAPE of the diagnosis input data models.[vii]

Cummings et al., 2002 performed a comparison of the predictive power of seven major risk adjustment systems that utilize administrative data.[63] After the models were calibrated, through regression to the population, R^2 values for concurrent models ranged from 0.24 (for Medicaid RX) to 0.47 (for DCGs). Because concurrent models estimate costs for treating known conditions, versus prospective models (such as ACGs) which use current data to predict cost outcomes for the next year, R^2 values were highest for concurrent models.[64]

Summary statistical performance measures, such as the R^2 value and c-statistics, are an indication of how well a risk adjustment model predicts the outcomes of interest or in discerning between patients without and without the outcome. R^2 value measures how well a risk adjustment system predicts a continuous outcome and gauges the extent to which the independent variables in the model explain variation in the dependent variable. While it is driven by the ratio of total variability in the outcome to a model-specific measure of the variability of actual values from model predictions, R^2 depends heavily on the features of the data set used and the amount of variation in both the dependent and independent variables.[65] Significant differences in reported R^2 value for different risk adjustment methodologies may merely reflect the difficulty in predicting outcomes in one particular database, rather than suggest substantive

---

[viivii] The MEDai model is used to forecast cost, inpatient stays, emergency room visits, prescription costs and savings, using medical and pharmacy claims, demographic information, lab results, and results of health risk assessments (HRAs). Member-level predictions are made using clinical risk factors, drug categories, age, gender, insurance type, and risk factors related to the timing and frequency of treatment.

differences between risk adjustment systems. It may, therefore, be misleading to rely exclusively on these indicators to select a risk adjustment model. Factors such as potential confounding between the independent variables and the dependent variable, inflation of the summary statistic by select variables, and data manipulation should all be taken into consideration.

Further, summary statistics do not specify the ability of risk adjustment models to distinguish outcomes among subgroups of patients. For example, risk adjustment controlling for ethnicity will not reveal how performance among two providers have similar or divergent trends for each particular ethnicity. This presents the case for introducing risk stratification as the first step of risk adjustment in cases where policy-sensitive patient characteristics (e.g., race, socioeconomic status) are important risk factors but may also reflect differences in the treatments patients receive.[2] Risk stratification examines performance within stratums of patients defined by defined characteristics. An ethnicity-stratified comparison, or comparing the outcome by ethnicity across the two providers, may be beneficial.

E. Maximizing the impact of risk adjustment

Acuity and adherence

Patient acuity can be interpreted as a measurement of the intensity of care required for a patient, by a provider. Generally, acuity categories range from low acuity for minimal care, to high acuity for intensive care. Little is known about the relationship between acuity and outcome; a standardized approach to measuring acuity is not available.[66] Some research suggests that the Medicare case mix index (CMI) can serve as a proxy for patient acuity.[67]

Accurate attribution of a patient to a physician or facility and a physician to a medical group are both critical to ensuring that physician and group scores are calculated and interpreted correctly.[68] Physician identifiers may not be available or accurate on all claims. AHRQ cites that three leading methods of assigning physicians to groups using claims data (Unique Physician Identifiers, Taxpayer Identification Numbers, and physician group roster) have inaccuracies of 10 percent or more.[68]

Attribution for members who visit multiple physicians throughout the year or are attributed to multiple providers due to multiple conditions

When care is provided by more than one provider, such as in episodes of care, the attribution of primary accountability for the patient and resources used is often called into question. No known national consensus guidelines for provider attribution are available. Various algorithms based on visit counts and payment amounts have been developed and applied, each adjusting for local market characteristics such as the availability of certain types of providers, level of community resources, or geographic isolation.[68]

Attribution error and bias

Accurate assessment of health care quality depends heavily on the accuracy of assigning responsibility to providers for their role in assuring that patients receive the necessary services and achieve desired health outcomes. In each case it must not only be ascertained that the provider's action or inaction *can cause* a particular outcome to occur, it must also be shown that the observed processes or outcomes are *attributable to* the provider.[2] Though the first condition must clearly be satisfied, the second condition may be complicated by instances in which the provider carries out the standard of care under the given circumstances and context, but the patient either chooses not to follow through with treatment or is prevented from obtaining treatment due to financial limitations, health insurance coverage, or personal

health beliefs. Or, perhaps the patient received the correct treatment, but the process was unsuccessfully documented in the health information systems. In such cases, an adverse outcome or missing record may nevertheless reflect poorly on the provider, even though he or she performed satisfactorily within his or her domain of influence.

As described, the patient's choice and circumstances may be confounding variables in the relationship between provider performance and patient outcome. Confounding in quality measurement remains a significant issue. When the association between two variables is influenced by a confounding variable, the confounder may be the cause of all or a portion of the observation link between the first two variables. Confounding variables may be due to patient-level characteristics, provider practice resources, health system policies, or the availability or accessibility of health care resources.[69]Adjustment for confounding variables using statistical modeling is a cornerstone of risk adjustment. Furthermore, providers may be sorted into subgroups by the characteristics of patients they serve or facilities in which they practice. Variation in the action or inaction of providers within subgroups tend be more consistent. This clustering of data may be overlooked by standard regression models, leading to possible bias in performance results. In such cases, hierarchical modeling is able to capture variation arising from differences between providers and subgroups of providers.[2]

Recent research has indicated that providers who practice in resource-limited environments and/or care for more complex patients are at higher risk of bias in performance measurement. Clinicians practicing in communities that have lower socioeconomic status, fewer public health resources, less-developed infrastructure tend to have less control over acquiring and directing the use of essential resources in care delivery, such as equipment, personnel, facilities, and information systems.[70] Further, community-level factors may influence the health and outcomes of local residents.[2] The myriad of potential factors adversely interacting with health outcomes can be difficult to capture and include in the risk adjustment process.[71]

Clinicians who practice in resource-limited environments tend to treat a higher proportion of complex patients.[71] Complex patients may exhibit greater severity of illness or a higher number of comorbidities. Alternatively, complexity may also arise from patient-level characteristics such as socioeconomic status, behavior, or health beliefs. Although risk adjustment accounts for severity of comorbidities, additional sources of data would be necessary to include a complete set of patient-level characteristics. However, even the most advanced systems for data recording and reporting may not accurately or reliably capture certain patient-level characteristics.[72] The growing widespread use of electronic health record (EHR) systems may complicate this issue further, as such programs may offer the clinician less flexibility in recording patient information that may not be required but may be helpful in risk adjustment, compared to traditional handwritten records.

Other Considerations

A. Administrative burden

The process of providing data for performance analysis, data collection, data mining, and tailoring of the risk adjustment model can all lead to increased administrative burden for personnel. Steps should be taken to minimize provider disruption and administrative burdens by streamlining the data collection system or allowing the provider to select a data uploading method most suitable to his or her needs. The risk adjustment methodology should leverage existing data sources and build upon current systems infrastructure.

B. Limitations of risk adjustment

The principal limitation to risk adjustment is that it can only account for measureable and reported risk factors, as previously discussed. Many patient risk factors, such as behavioral or social factors and physiological conditions, may be difficult to measure or are reported on an inconsistent basis. Despite these limitations, outcome measures at the hospital-level are relatively reliable due to high between-hospital variation and randomly distributed unmeasured risk factors across hospitals.[73-74] At the physician-level, the limitations may have a greater impact due to the clustering of certain types of patients in certain physician practices.[68] Although hierarchical modeling may be an antidote to this dilemma, concerns about the reliability of quality measurement in such settings remain.

The validity and reliability of quality measures continues to be a significant concern among health care providers, insurers, patients, and other stakeholders. While reliance on evidence-based quality measures and accurate and complete data sources are commonly cited essential components, the physician-level reliability of quality measures is another key consideration, particularly when incentives are tied to quality measurement. Precisely, physician-level reliability refers to a quality measure's ability to discern an individual provider's performance from the performance of physicians overall and depends on having a sufficient number of patients eligible for inclusion in a given quality measure and variation across physicians on that quality measure.[75-76]

Reliability of performance measurement

Scholle et al., 2008 investigated physician-level reliability of twenty-seven commonly used performance measures that are based on administrative data, including many that overlap with the Medicaid Adult Core Set. Although there was no attempt to risk adjust for differences in case mix or severity across physicians in this study, its findings on adequate sample size are nonetheless relevant to the current discussion. The study relied on a large dataset that combined patient-level administrative data from across nine large health plans to measure performance of primary care physicians; the dataset was chosen to model a typical data source used by individual health plans to profile physician performance. The researchers define high reliability in physician quality measurement as having comparatively high physician-to-physician variance or when there is low measurement error in the measurement of individual physician performance generally due to large sample sizes. By their assessment, the denominator of eligible patients for an individual physician from a single dataset is often too small for the results to be reliable.[75] Most quality measures require at least 50 patients eligible for a quality measure to be included in the sample size to obtain a reliable estimate of individual physician performance. Large sample sizes per physician are needed to achieve optimal reliability when assessing performance at the individual-level.

Data completeness and reliability can be improved by using composite measures, group-level reporting, combining multiple years of data, or combining health care claims data from multiple carriers.[68] An example of a commonly used group-level technique is physician group-level reporting. However, potential implementation issues that arise include difficulty in assigning physicians who belong to multiple groups or fluid group structures, exclusion of providers in solo or small-group practices, or poor attribution of patients to physician groups. Alternatively, combining multiple years of data may improve the reliability of physician-level reporting, but the ability of historical data to make inferences about current or future performance diminishes when older data is used. Many collaboratives across the country, such as chartered value exchanges (CVEs) and the Wisconsin Collaborative for Healthcare Quality are using combined health care claims data from multiple payers, with successful results.[68]

Indirectly standardized outcome ratios

Risk-adjusted outcome measures generally compare the actual outcomes of a specific set of patients treated by a provider or facility with their expected outcomes had they been treated by an average provider or facility. Therefore, each provider or facility is compared with a hypothetical counterfactual representing the average provider or facility treating the same group of patients, as opposed to comparing to an actual peer treating a different set of patients. Because of this, the AHRQ does not recommend numerically ranking hospitals based on risk-adjusted outcomes.[68] Instead, it recommends the placement of hospitals in 3-5 "bins" based on statistical criteria and ordered alphabetically or geographically within the bins.

Outlier data is treated in a variety of ways in risk adjustment. While excluding outliers is a common approach, other statistical approaches may be preferable. For example, it is possible to truncate outliers to reduce their effects on the analyses. This procedure is referred to as winsorizing, whereby a lesser weight is assigned to the outlier or its value is modified so it resembles other sample values. A potential downside is that winsorizing and eliminating outliers may introduce statistical bias and undervalue the outlier. On the other hand, including the outlier without modification may overvalue the outlier, leading the estimate to vary significantly from the true population value.[77]

Recommendations

The methodology selected to risk adjust quality measures will undergo significant scrutiny by stakeholders. Therefore, the methodology must strike a balance between quantitative rigor, ease of explanation, clinical plausibility, and conceptual meaningfulness between outcomes measures and risk factors. We recommend that the following factors be taken into consideration and offer recommendations for each factor.

A. Risk adjustment grouper

The selected methodology must accurately measure the health status of each individual and must be validated in the literature. It is also important that the methodology is widely used and understood in the community. The methodology should also utilize a categorical approach to assigning a member's health status, or a categorical structure should be overlaid on an additive model. A categorical structure allows for the measurement of adherence rates by acuity group and controls for differences between provider risk distributions versus the population or comparison group.

Recommendation

In evaluating these criteria, we recommend that DHS employ the ACG grouper in risk adjusting quality measures. Its categorical structure, public acceptance, and statistical performance all make it an excellent choice. In addition to the ability to use the ACG or Resource Utilization Bands (RUBs) risk groups to stratify the population, Expanded Diagnosis Cluster (EDCs) may also prove valuable in identifying populations with chronic conditions of interest.

B. Stratification

In addition to measuring the acuity of a member, the methodology should also evaluate whether socio-economic and other member characteristics are helpful in stratifying the population to improve the accuracy of the results. Member characteristics such as race/ethnicity, language, household income and education may all be related to variation in adherence rates, and these factors are not factored into a member's acuity classification. Region of residence and the local health care service system is also likely to be affiliated with adherence rates.

Recommendation

Lewin will work with DHS to develop a list of member characteristics to evaluate for subpopulation stratification. Lewin will then evaluate the impact of the selected characteristics on understanding variations in adherence rates for each measure. Lewin will review the results of this analysis with DHS to determine those characteristics that should be utilized to stratify the population. The selection of measures should take into consideration the statistical significance of a measure and the likely public acceptance of the characteristic. Statistical significance will be determined based on the model characteristics and sample population. For example, in the ERG model to assess the value of the risk adjustment on the basis of acuity level, a logistic regression model of acuity level on adherence is created, and the statistical significance of the association between member acuity and adherence is assessed using the p-value. If the p-value for acuity level is not statistically significant, the impact of risk adjustment was deemed not significant overall. If the adherence rates did not increase with the acuity level, the impact was also determined to be not significant.

In the example risk adjustment analysis for a sample Medicaid population presented in Appendix B, statistical significance was achieved if the p-value was below 0.05, indicating a small likelihood that the association between acuity level and adherence arose due to chance. We will use the same statistical benchmark (0.05) to identify significant characteristics in the Minnesota data.

C. Selection of measures

Lewin has risk adjusted the AQM measures for several states. Lewin has also explored the impact of risk adjustment on other quality measures for chronic conditions with high prevalence rates. Our evaluation of the results of these analyses has revealed that some measures are better choices for risk adjustment than others.

Although a variety of factors will be considered and tested, our analysis has found that two factors, in particular, should be taken into consideration in selecting measures for risk adjustment:

Member Distribution by Risk Group: Measures that are targeted for members with significant chronic conditions may have a very narrow distribution across acuity categories. Therefore, the application of risk adjustment offers little additional insight into a provider's performance. Other measures that are applicable to a broader population often have more dispersion across risk groups and offer more insights into a provider's performance.

Variations in Adherence Rate by Risk Group: Many measures have a broad distribution across risk groups, but show little variation in adherence rates from group to group. The application of risk adjustment in this instance may also offer little additional insight into a provider's performance. Regardless of the risk distribution of a provider's members, the expected adherence rate from provider to provider may show little variation. Other measures with larger variations in adherence rates across risk groups will provide better insights into a provider's performance.

After selecting measures for risk adjustment, an analysis of possible risk factors to include for each selected measure will be performed. For example, clinical evidence and literature suggest that certain socio-demographic risk factors, such as race and age, are highly associated with breast cancer prevalence.[78]

Recommendation

DHS will select quality measures that it seeks to risk adjust. Prior to finalizing the list of measures, the impact of risk adjustment on each measure should be evaluated. Lewin will work with DHS to evaluate the distribution of members identified for each measure across acuity groups by running Medicaid claims data (including eligibility, provider, institutional, professional, and pharmacy data) through EBM Connect to generate performance result statistics. If risk adjustment is found to provide little insight into the adherence of the population with a measure, either because of the acuity distribution or fluctuation in adherence by acuity group, DHS should also evaluate the impact of stratification by member characteristic. If neither approach is found to offer insights into fluctuation in member adherence, DHS may want to consider revising the list of selected measures. The statistical criteria for determining whether a measure would benefit from risk adjustment will depend on the p values when the measure is stratified by acuity group.

D. Provider performance

Providers' performance should be measured in comparison to their peers. For example, comparisons between pediatricians and gerontologists may not be meaningful and may come under criticism from stakeholders. The measures where pediatricians have sufficient number of patients that meet the selection criteria to provide credible results will differ from the measures with credible numbers of patients for gerontologists Our prior experience indicates that having 30-50 members per measure is generally sufficient for meaningful comparison. The provider classification system must choose a proper balance between defining groups of homogeneous providers and ensuring that each group has enough providers in order to provide meaningful comparisons. Peer group sample size criterion will be determined based on characteristics and availability of data.

A provider classification methodology will be used to assign providers to a peer group. The provider classification system must choose a proper balance between defining groups of homogeneous providers, in terms of their clinical practice and patient panel, and ensuring that each group has enough providers in order to provide meaningful comparisons.

Some groups to consider in developing the classification system are:

- Pediatricians
- General Practitioners
- Internal Medicine
- Specialist
- Community Based Clinics

Provider performance across provider peer groups may also be useful in circumstances in which different types of providers may treat members that qualify for the same measure. Providers may apply dissimilar treatment methods depending upon the focus of their practice. Endocrinologists, for example, may focus on treatments related to the care of a member's diabetes and place less focus on cervical and breast cancer screenings.

Recommendation

Lewin and DHS should develop a provider classification methodology to evaluate the performance of providers relative to their peer group. This approach will provide an additional level of stratification in the methodology and will likely improve stakeholder acceptance. The provider classification system must choose a proper balance between defining groups of homogeneous providers and ensuring that each group has enough providers in order to provide meaningful comparisons. The expected adherence rate of a provider's panel will be developed based upon the average adherence rate for their peer group. The performance of a pediatrician, for example, would then be compared to the performance of the average pediatrician and not the average provider.

Phase Two - Analysis

Phase two of this project was the analysis that consisted of testing and implementing the risk adjustment methods. As previously stated, the Medicaid Adult Core Set measures were the focus of this risk adjustment project. The Affordable Care Act (ACA) requires the U.S. Department of Health and Human Services (HHS) to identify and publish quality measures to help monitor the quality of care for Medicaid-eligible adults. The initial core set of 26 adult health care quality measures were published by HHS in January 2012; more information and background on the Medicaid Adult Core Set can be found on CMS' website (www.medicaid.gov). DHS sought to improve the use of these quality measures and facilitate more accurate comparisons across groups such as MCOs. The first step was to select a subset of the measures on which to examine the impact of risk adjustment. DHS grouped the measures into related clinical areas and selected 19 measures of interest. The measures encompassed the following groups: preventive women's health, chronic, mental health, behavioral, chronic hospitalization, and treatment measures. Table 9 summarizes the selected quality measures. See Appendix A for additional information about these measures and quality measurement.

Table 9. Selected quality measures.

Category and Acronym	Description
Preventive Women's Health	
BCS-AD	Breast cancer screening
CCS-AD	Cervical cancer screening
CHL-AD	Chlamydia screening in women
Chronic	
MPM-AD-R1	Annual monitoring for enrollees on angiotensin converting enzyme (ACE) inhibitors or angiotensin receptor blockers (ARB)
MPM-AD-R2	Annual monitoring for enrollees on digoxin
MPM-AD-R3	Annual monitoring for enrollees on diuretics
MPM-AD-R4	Annual monitoring for enrollees on anticonvulsants
HA1C-AD	Comprehensive diabetes care: Hemoglobin A1c testing
Mental Health	
FUH-AD-7	Follow-up after hospitalization for mental illness (7-day)
FUH-AD-30	Follow-up after hospitalization for mental illness (30-day)
SAA-AD	Adherence to antipsychotics for individuals with schizophrenia
AMM-AD_acute	Antidepressant medication management (acute phase)
AMM-AD_cont	Antidepressant medication management (continuation phase)
Behavioral	
IET-AD-14	Initiation and engagement of alcohol and drug dependence treatment (14-day)
IET-AD-30	Initiation and engagement of alcohol and drug dependence treatment (30-day)
Chronic Hospitalization	
PQI01-AD	Diabetes short-term complications admission rate
PQI05-AD	Chronic obstructive pulmonary disease (COPD) or asthma in older adults admission rate
PQI08-AD	Heart failure admission rate
Treatment	
PPC-AD	Postpartum care rate

Next, decisions were made regarding characteristics that would potentially be included in the risk adjustment. Of specific interest were patient-related characteristics that could influence a person's outcome on a given quality measure, but are beyond the control of the MCO. This included characteristics such as health risk and age, which are commonly included in risk adjustment analyses. In addition, DHS was particularly interested in other factors such as sociodemographic characteristics. Accordingly, we identified characteristics suggested by the National Quality Forum (NQF) as potential sociodemographic characteristics to explore when risk adjusting quality measures.

Once these decisions regarding quality measures and potential characteristics were made, the analysis could proceed. The first steps encompassed DHS providing the necessary data, assessing data quality, and computing the quality measures. Specifically, data from nine Minnesota MCOs and FFS data was provided for this analysis. The data encompassed Medicaid claims, enrollment, and provider data in addition to measures produced by the Johns Hopkins Adjusted Clinical Groups® (ACG®) system.[79] ACGs are categories (defined by morbidity, age, and gender) that group people with similar health or illness burden.[80] The ACGs were the source for our measure of health risk, resource utilization bands (RUBs). The RUBs combine people across the ACG clinical categories into groups expected to have similar healthcare needs and resource utilization.[80] The ACG software assigns people into six RUB categories based on demographic (i.e., age and gender) and clinical information such as health care claims diagnosis codes: non-user, healthy, low morbidity, moderate morbidity, high morbidity, and very high morbidity. The Lewin Group used the DHS data combined with the EBM Connect® software[81] to derive the performance results for the quality measures for Minnesota Health Care Program enrollees; the 2013 measurement year was used for this analysis.

Next, statistical methods were employed to guide the selection of suitable characteristics to include in the risk adjustment. This process was guided, in part, by statistical criteria that assessed the extent candidate characteristics were related with each of the Medicaid Adult Core Set measures. The statistics effectively provide suggestions for which factors should be included. In addition, non-statistical considerations also guided this process; DHS identified select characteristics for inclusion based on their subject matter expertise and policy implications. A statistical model was created for each of the examined Medicaid Adult Core Set measures, which was comprised of the list of characteristics selected from this process.

In order to study the impact of risk adjusting the Medicaid Adult Core Set quality measures, we applied the statistical models to risk adjust the results for each of DHS' contracted MCOs. This allowed us to examine how accounting for the patient-related characteristics in the models altered the MCOs adherence rates on these quality measures given they each have different mixes of patients.

For more details on the methodology used in the analysis please refer to Appendix A.

Analysis Findings

A. Risk Adjustment Models

Almost all of the quality measures we examined (17/19 or 89%) were well suited for risk adjustment. The measures were related to patient characteristics that differed across MCOs. Lewin identified a variety of patient-related characteristics that were related to outcomes on the quality measures. Table 10 summarizes the variables retained in the model for each of the quality measures and indicates whether those variables were statistically significant. The models included a total of twelve factors encompassing six clinical and six sociodemographic characteristics. The clinical factors encompassed a member's overall health risk (i.e., Resource Utilization Bands or RUB), whether the member had a developmental disability, was enrolled in Medicaid due to a disability, was frail, had a mental health condition, or was identified as having a substance abuse issue. The sociodemographic characteristics encompassed a member's age, gender, education, language, race and ethnicity, and whether the member lived in a metropolitan county. These characteristics, health risk, for example, are outside the control of MCOs and the patient mix among these characteristics differed across MCOs. One MCO had roughly 20% of people in the very high morbidity category (i.e., meaning worse health and higher risk) whereas other MCOs had as low as roughly 3%. Table 22 in Appendix C provides the frequency distribution for these variables for each of the MCOs and the FFS population. Therefore, this table exhibits how each MCO's population potentially differed among these characteristics.

In addition to which factors were included in the risk adjustment, we also wanted to gauge which of those characteristics had larger influence. Table 11 summarizes the magnitude of influence of the characteristics. Health risk and age were consistently influential factors. This is supported by both the number of statistically significant relationships and the magnitude of influence of these characteristics. The measures of health risk (i.e., RUBs) and age were the characteristics with the most statistically significant relationships across the quality measures (see Table 10). In addition, the amount of influence from these characteristics was commonly larger compared to the other characteristics (see Table 11 or Appendix C for more detailed statistics). Health risk also had the most consistent pattern of influence across the quality measures. In general, as health risk increased (i.e., higher morbidity) people were more likely to satisfy the measures (e.g., be a "yes" such as having had a mammogram on the breast cancer screening measure).

The remaining sociodemographic and clinical characteristics were also important. These characteristics each commonly had a number of statistically significant relationships across the quality measures (see Table 10). The amount of influence these factors had on the quality measures was mixed and had both positive (e.g., increasing the likelihood on a measure such as having a mammogram) and negative associations. In general, these characteristics exhibited smaller influence relative to health risk and age (see Table 11). Even so, sometimes these characteristics had larger influence on select quality measures.

Table 10 also exhibits where variables were included for some measures when the relationship was not statistically significant. To recap, the variable selection process encompassed both statistical criteria and non-statistical decisions. For example, of particular policy interest to DHS was an examination of sociodemographic characteristics. As demonstrated in Table 10, these characteristics commonly had significant relationships with multiple, but not all quality measures. Given the overall statistical merits and policy importance of these factors, these characteristics were included in several models for reasons other than statistical criteria.

Risk adjustment models were not feasible for two of the examined quality measures due to low sample size in these measures. The first was the measure assessing annual monitoring for patients on persistent

medications, "Annual monitoring for enrollees on digoxin" (MPM-AD-R2). The second was the chronic hospitalization measure assessing the congestive heart failure admission rate (PQI08-AD). Consequently, these measures show no selected variables in Table 10 (i.e., all cells in the column are dashes).

The detailed statistical results for the models for each quality measure are available in Appendix C.

Table 10. Summary of final models.

Variables	Measure Category and Acronym																		
	Preventive Women's Health			Chronic					Mental Health					Behavioral		Chronic Hospitalization			Treatment
	BCS-AD	CCS-AD	CHL-AD	MPM-AD-R1	MPM-AD-R2	MPM-AD-R3	MPM-AD-R4	HA1C-AD	FUH-AD-7	FUH-AD-30	SAA-AD	AMM-AD_acute	AMM-AD_cont	IET-AD-14	IET-AD-30	PQI01-AD	PQI05-AD	PQI08-AD	PPC-AD
Clinical																			
Developmental Disability	Y	Y*	Y*	Y	--	Y	Y*	Y	Y	Y*	Y*	Y*	Y*	Y*	Y			--	
Disability	Y*	Y*	Y*	Y*	--	Y	Y	Y*	Y	Y	Y*	Y	Y	Y	Y		Y*	--	
Frailty	Y*			Y	--	Y*	Y*	Y	Y	Y	Y*	Y*	Y*	Y*	Y*			--	
Mental Health	Y	Y	Y*	Y	--	Y	Y	Y*						Y*	Y			--	
Resource Utilization Bands (RUB)	Y*	Y*	Y*	Y*	--	Y*	Y*	Y*	Y*	Y*	Y	Y*	Y*	Y*	Y*	Y*	Y*	--	Y*
Substance Abuse	Y	Y*	Y*	Y	--	Y	Y*	Y*	Y	Y	Y*	Y*	Y*					--	
Sociodemographic																			
Age	Y*	Y*		Y*	--	Y*	Y	Y*	Y	Y	Y*	Y*	Y*	Y*	Y*	Y*	Y*	--	Y
Education	Y*	Y*	Y	Y	--	Y*	Y	Y	Y*	Y*	Y	Y*	Y*	Y	Y	Y	Y*	--	Y*
Gender				Y	--	Y	Y	Y	Y*	Y*	Y*	Y*	Y*	Y	Y	Y*	Y*	--	
Language	Y*	Y*	Y*	Y*	--	Y	Y	Y*	Y	Y	Y*	Y	Y	Y	Y*	Y*	Y*	--	Y*
Metropolitan County	Y	Y*	Y*	Y*	--	Y*	Y*	Y*	Y*	Y*	Y	Y	Y	Y	Y	Y	Y	--	Y*
Race / Ethnicity	Y	Y*	Y*	Y	--	Y	Y	Y*	Y	Y	Y*	Y*	Y*	Y*	Y*	Y*	Y	--	Y*

Notes: The measures are listed by acronym in this table for brevity. Please reference Table 9 for measure names and descriptions. The Y symbol in a cell denotes the variable was included in a given model. The asterisk symbol (*) denotes the variable was statistically significant ($p < 0.05$). The two columns where the cells contain dashes (i.e., --) identify the two measures that were not able to be risk adjusted (e.g., due to sample size and model convergence issues).

Table 11. Summary of the influence of individual characteristics across the quality measures.

Characteristic	Level	Negative Association Average Odds Ratio (N)	Positive Association Average Odds Ratio (N)
Developmental Disability	No vs. Yes	0.79 (11)	1.53 (3)
Disability	No vs. Yes	0.79 (6)	1.40 (9)
Frailty	No vs. Yes	0.80 (10)	1.28 (2)
Mental Health	No vs. Yes	0.79 (5)	1.09 (4)
Resource Utilization Bands (RUB)	Non-user vs. Very High	0.06 (11)	1.63 (3)
Resource Utilization Bands (RUB)	Healthy User vs. Very High	0.33 (17)	-- (0)
Resource Utilization Bands (RUB)	Low Morbidity vs. Very High	0.36 (16)	1.06 (1)
Resource Utilization Bands (RUB)	Moderate Morbidity vs. Very High	0.55 (15)	1.17 (2)
Resource Utilization Bands (RUB)	High Morbidity vs. Very High	0.65 (14)	1.27 (3)
Substance Abuse	No vs. Yes	0.89 (9)	1.28 (3)
Age	45-64 vs 65+	-- (0)	1.65 (1)
Age	45-64 vs. Under 45	0.75 (8)	1.69 (7)
Age	65+ vs. Under 45	0.50 (5)	2.60 (7)
Gender	Female vs. Male	0.87 (7)	1.21 (6)
Education	Unknown vs. More than High School	0.86 (6)	1.29 (11)
Education	Less than High School vs. More than High School	0.84 (9)	1.19 (8)
Education	High School Graduate vs. More than High School	0.88 (11)	1.21 (6)
Language	Non-English vs. English	0.74 (8)	1.28 (9)
Metropolitan County	No vs. Yes	0.86 (13)	1.11 (4)
Race / Ethnicity	Non-white vs. White	0.75 (10)	1.21 (7)

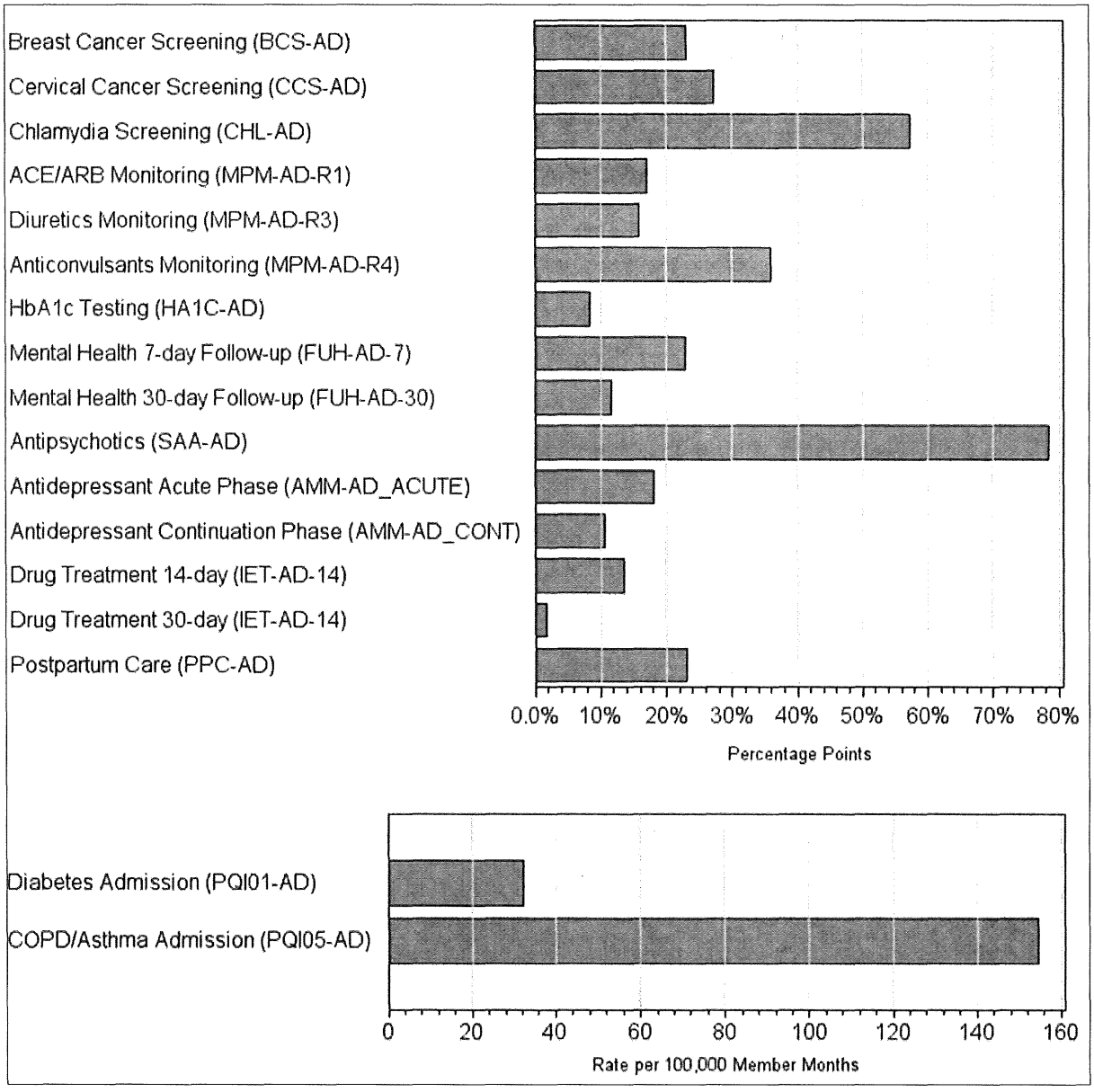
Notes: The N is the count of quality measures. The odds ratio represents the size of influence, on average, across the quality measures. Values less than 1.0 denote a negative association (e.g., less likely to satisfy a measure such as receiving a mammogram to screen for breast cancer); a value closer to zero indicates a larger influence than a value closer to 1.0. Values greater than 1.0 denote a positive association (e.g., more likely to satisfy a measure); a value greater than and further from 1.0 indicates a larger influence than a value closer to 1.0.

B. Risk Adjusted Rates

Overall, the influence of risk adjustment on the quality measures was mixed with some MCO adjusted performance trending upward from its unadjusted figure while other MCOs experienced a reduction in their performance estimate after applying risk adjustment to the respective quality measures. Across the 17 risk adjusted measures, over 50% of the MCOs' rates improved for the majority of measures (12 measures or 71%). Two of the preventive women's health measures showed the least improvement

following risk adjustment. For the cervical cancer screening measure, only 22% of MCOs' (2/9) rates improved. For the breast cancer screening measure, only 33% of MCO's (3/9) rates improved. Across the 9 MCOs, most MCOs (5/9 or 56%) saw rates improve on more than 50% of the measures following risk adjustment. The largest number of improvements observed by an MCO was 76% (13/17) of measures and the lowest was 12% (2/17) measures.

Figure 1. Aggregate impact of risk adjustment across MCOs, by quality measure.



Note: These values are the absolute value of the difference between the adjusted and unadjusted rates summed across the MCOs.

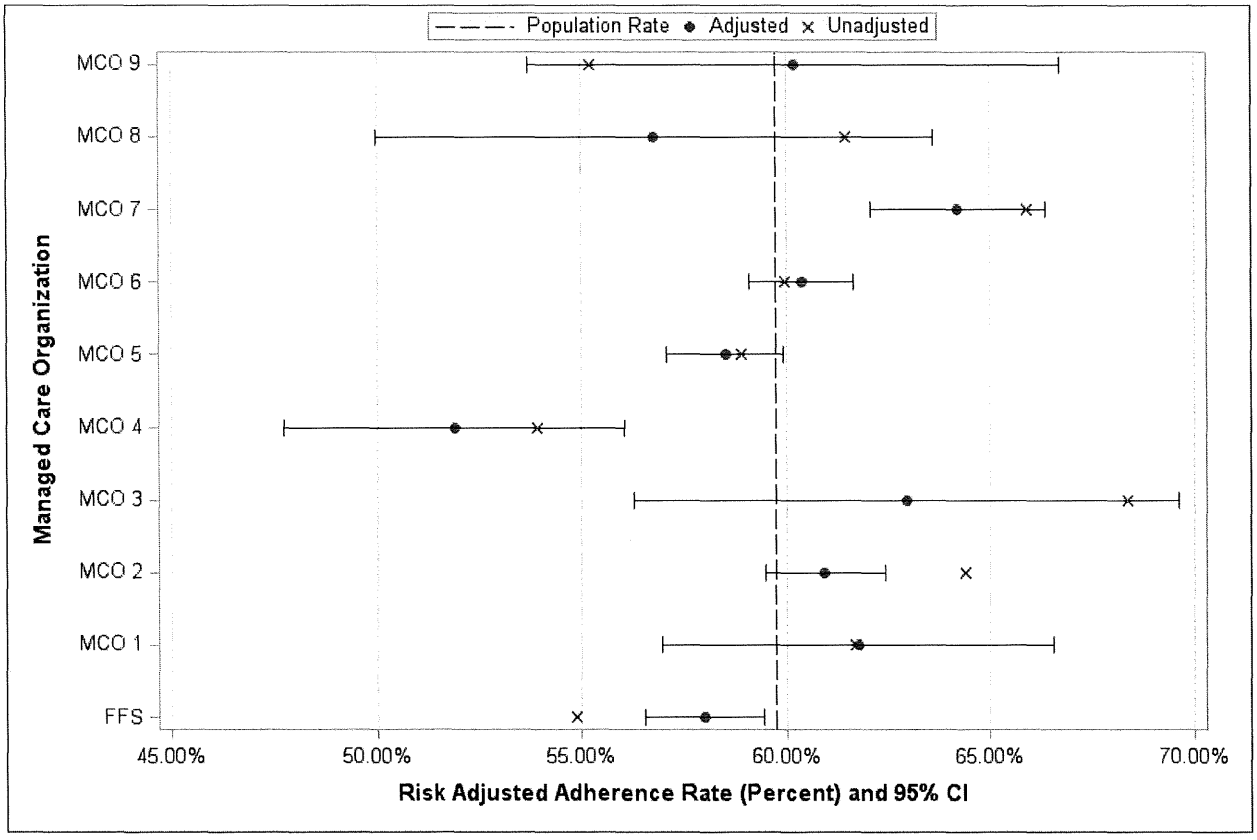
Figure 1 illustrates the extent risk adjustment impacted the rates for each quality measure. The values are the absolute value of the difference between the unadjusted and adjusted rates, summed across the MCOs. For example, among the non-PQI measures, if the change from the unadjusted rate to the adjusted rate was 40% to 45% (i.e., 5% difference) for one MCO and 65% to 62% (i.e., -3% difference and absolute

value of 3%) for a second MCO, the summed aggregate impact for these two MCOs would be 8% (i.e., 5% + 3%). The calculation was the same for the PQI measures. For example, if the change was 10/100,000 member months to 6/100,000 members months for an MCO, the absolute value of the change in the rate is 4/100,000 member months. Among the non-PQI measures, the antipsychotic measure related to adherence among individuals with schizophrenia had the largest aggregate change in rates. The 30-day drug treatment measure had the smallest change. Among the PQI measures, the COPD/asthma admissions measure had a larger change. Even so, the impact of risk adjustment is based on factors such as the characteristics included in the risk adjustment models, the relationships between those characteristics and each measure, and the patient mix of each MCO. Therefore, as these factors change over time (e.g., patient mix), the relative impact of risk adjustment on each measure can also change.

Due to the volume of examined quality measures, we present the results for only a single measure in the main report. We selected the breast cancer screening measure because it is easily understood. Results for the remaining measures can be found in Appendix C.

Figure 2 is a plot of the results for the Preventive Women's Health measure, breast cancer screening. This plot shows each managed care organization's unadjusted and adjusted performance rates and the confidence intervals. Displaying results in this manner allows us to compare each MCO to the statewide rate, assess if any observed differences are meaningful, and gauge the impact of risk adjustment on that organization.

Figure 2. Risk adjusted rates for the breast cancer screening measure (BCS-AD), by MCO.



Closer examination of these plots yields several key pieces of information. The vertical dashed line represents the overall population rate on this measure (59.74%). The 95% confidence interval represents

the upper and lower limit of the estimated rate for each MCO and allows us to assess if departures from the population rate are meaningful once statistical fluctuation has been taken into account. For example, considering the FFS group, the confidence interval does not overlap with the overall population rate. Consequently, it can be stated that the adherence rate on this breast cancer measure for the FFS group was almost certainly lower than the statewide rate. While we selected only the breast cancer screening measure for illustration in the main report, this measure generally reflects the results for the remaining measures. For most of the quality measures, the confidence intervals for many MCOs overlapped with the population rate.

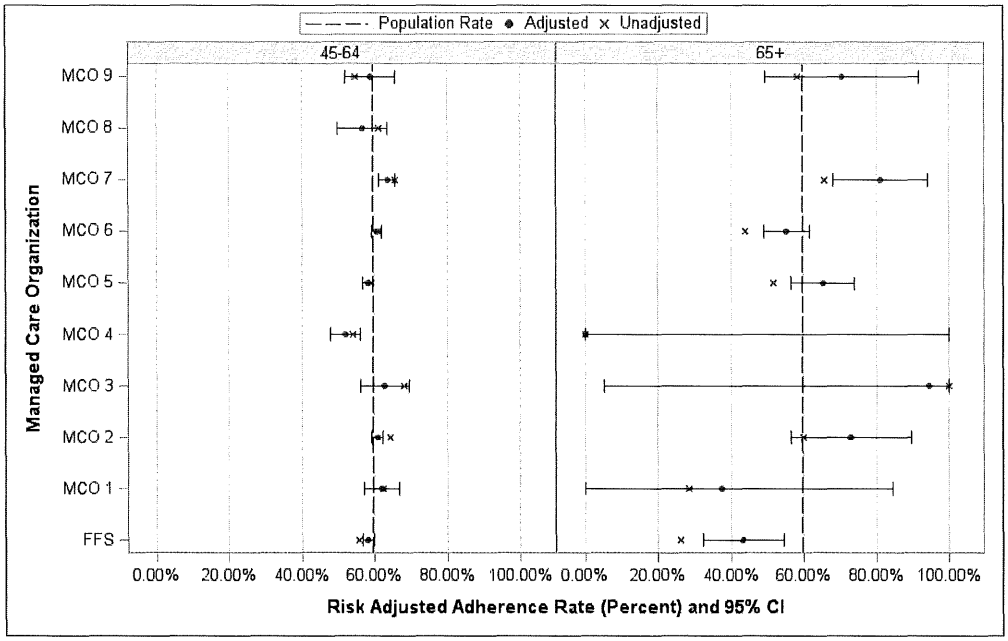
In addition to the figures, tables containing the values and statistics related to the risk adjusted rates can also be found in Appendix C.

Stratified Risk Adjusted Rates

When sociodemographic characteristics are included in the risk adjustment of quality measures, it has been recommended to also stratify results.[86] The rationale is that presenting the stratified results will more clearly highlight any potential disparities. In order to keep this report to a manageable length, only the age stratified results for the breast cancer screening measure are presented as an example in Figure 3. Please see Appendix C for additional stratification results and tables.

Consider the age stratifications for the FFS group as an example. We saw in Figure 2 that, although the risk adjusted rate (58.00%) improved relative to the unadjusted rate (54.86%), the adjusted rate was still slightly lower than the overall population rate (59.74%). In the stratified results shown in Figure 3, we see this difference is mostly attributable to the wider gap one can observe for the older age group (i.e., ages 65+). While the risk adjusted rates improved in both age groups compared to the unadjusted rates, the older age group had a lower risk adjusted rate (43.15%) compared to the younger age group (58.34%). Similarly, we see across the MCOs that risk adjustment can have a different impact within subgroups of each MCO's population such as age groups.

Figure 3. Risk adjusted rates for the breast cancer screening measure (BCS-AD), stratified by age group and MCO.



Conclusions and Recommendations

These analyses empirically demonstrate the influence of an array of clinical and sociodemographic characteristics on adherence rates among the Initial Core Set of Health Care Quality Measures for Medicaid-Eligible Adults. This is important because (1) MCOs commonly have different mixes of patients across the characteristics we found to influence these quality measures and (2) these patient-related attributes are beyond the control of MCOs. Consequently, not accounting for these patient-related differences could result in imperfect comparisons when contrasting groups such as MCOs. Ultimately, these analyses exhibit the need and appropriateness of risk adjusting quality measures when aiming to compare subgroups such as MCOs.

These analyses demonstrated that the impact of risk adjustment can differ within subgroups of sociodemographic characteristics such as age groups within an MCO. Therefore, these results align with recommendations by the NQF [82] and support the notion that the results of risk adjustment that include sociodemographic characteristics should be stratified. Stratifying by the sociodemographic characteristics presents the results across the individual subgroups thereby helping to prevent the risk adjustment from masking any potential disparities.

Regarding the characteristics that should be accounted for in the risk adjustment, the most consistent and influential characteristics were health risk and age. While the size of influence and importance was mixed, these analyses also demonstrated that other sociodemographic characteristics can impact quality measures. In addition to age, this included a member's education, gender, language, race and ethnicity, and whether they lived in a metropolitan county. While these characteristics generally had a smaller influence (i.e., relative to health risk and age), there were notable exceptions. These results suggest that sociodemographic characteristics should be considered when exploring risk adjustment of quality measures.

We observed similar results amongst the remaining clinical characteristics chosen for risk adjustment consideration. While the size of influence and importance was mixed, these analyses demonstrated clinical characteristics beyond a general measure of health risk can impact quality measures. This included whether a member had a developmental disability, was enrolled in Medicaid due to a disability, was frail, had a mental health condition, or was identified with a substance abuse issue. Therefore, these results suggest that targeted clinical characteristics of interest that might not be fully captured in a health risk measure could also be considered when exploring risk adjustment of quality measures.

Given the demonstrated relationships of the patient-related characteristics with the quality measures and that these characteristics differed across MCOs, this risk adjustment approach is recommended for Minnesota DHS to allow more accurate comparisons of MCOs. Moreover, this approach is applicable to and recommended for other states with similar aims of comparing quality measure results across groups such as MCOs. MCOs in other states undoubtedly have different patient mixes and would likely exhibit relationships between patient-related characteristics and quality measures. Therefore, this risk adjustment approach is recommended to account for those characteristics and afford more accurate comparisons.

Appendix A

This appendix provides methodologic detail encompassing the examined quality measures, data sources, examined variables, and statistical methods.

A. Quality Measures

DHS seeks to develop a risk adjustment methodology to enhance the use of the Initial Core Set of Health Care Quality Measures for Medicaid-Eligible Adults and enable more accurate comparison between managed care organizations (MCOs) and accountable care organizations (ACOs). Accordingly, we explored risk adjusting select quality measures from the Medicaid Adult Core Set; more information is available on CMS' website (www.medicaid.gov). We used the EBM Connect® software[81] to derive the performance results for these measures for Minnesota Health Care Program enrollees for the 2013 measurement year. Table 12 summarizes the individual measures DHS prioritized for exploration as part of this project. DHS grouped these measures into related clinical areas (e.g., Preventive Women's Health), which aided organization of the results and also allowed examination of similar patterns within groups. The groups allowed us to assess the extent specific variables were consistently more important for some clinical areas. For example, a given variable might be important for Women's Preventive Health measures, while less important for another group such as the Mental Health measures.

Table 12. List of quality measures examined.

Category and Acronym	Name	Description
Preventive Women's Health		
BCS-AD	Breast cancer screening	Patient(s) 52 - 74 years of age that had a screening mammogram in last 27 reported months
CCS-AD	Cervical cancer screening	Women that had appropriate screening for cervical cancer
CHL-AD	Chlamydia screening in women	Patient(s) 21 - 24 years of age that had a chlamydia screening test in last 12 reported months
Chronic		
MPM-AD-R1	Annual monitoring for patients on persistent medications - Annual monitoring for enrollees on angiotensin converting enzyme (ACE) inhibitors or angiotensin receptor blockers (ARB)	Adult patients persistently taking Angiotensin converting enzyme (ACE) inhibitors or Angiotensin receptor blockers (ARB) who received a serum potassium test AND either a serum creatinine or blood urea nitrogen test within the last 12 reported months (HEDIS criteria)
MPM-AD-R2	Annual monitoring for patients on persistent medications - Annual monitoring for enrollees on digoxin	Adult patients persistently taking digoxin containing medication who received a serum potassium test AND either a serum creatinine or a blood urea nitrogen test within the last 12 reported months (HEDIS criteria)
MPM-AD-R3	Annual monitoring for patients on persistent medications - Annual monitoring for enrollees on diuretics	Adult patients persistently taking diuretic containing medication who received a serum potassium test AND either a serum creatinine or a blood urea nitrogen test within the last 12 reported months (HEDIS criteria)

Category and Acronym	Name	Description
MPM-AD-R4	Annual monitoring for patients on persistent medications – Annual monitoring for enrollees on anticonvulsants	Adult patient(s) persistently taking anticonvulsant medication(s) who received drug serum concentration test for anticonvulsant medication(s) within the last 12 reported months (HEDIS criteria)
HA1C-AD	Comprehensive diabetes care: Hemoglobin A1c testing	Patient(s) 18 - 75 years of age that had a HbA1c test in last 12 reported months
Mental Health		
FUH-AD-7	Follow-up after hospitalization for mental illness	Patient(s) hospitalized for mental illness that had a follow-up encounter with a mental health practitioner within 7 days after discharge
FUH-AD-30	Follow-up after hospitalization for mental illness	Patient(s) hospitalized for mental illness that had a follow-up encounter with a mental health practitioner within 30 days after discharge
SAA-AD	Adherence to antipsychotics for individuals with schizophrenia	Patient(s) with schizophrenia who remained on antipsychotic medication for at least 80% of their treatment period
AMM-AD_acute	Antidepressant medication management	Patient(s) with major depression who start an antidepressant medication that remained on treatment for at least 12 week (acute phase treatment)
AMM-AD_cont	Antidepressant medication management	Patient(s) with a major depression who start an antidepressant medication that remained on treatment for at least 6 months (continuation phase treatment)
Behavioral		
IET-AD-14	Initiation and engagement of alcohol and other drug dependence treatment	Patient(s) with a new episode of alcohol and other drug dependence (AOD) who initiated treatment within 14 days of the diagnosis
IET-AD-30	Initiation and engagement of alcohol and other drug dependence treatment	Patient(s) with a new episode of alcohol and other drug dependence (AOD) who initiated treatment and had two or more follow-up visits within 30 days of the initiation visit (i.e. engaged in AOD treatment)
Chronic Hospitalization		
PQI01-AD	Diabetes short-term complications admission rate	Patient(s) hospitalized for short-term complications of diabetes
PQI05-AD	Chronic obstructive pulmonary disease (COPD) or asthma in older adults admission rate	Patient(s) hospitalized for chronic obstructive pulmonary disease (COPD) or asthma
PQI08-AD	Heart failure admission rate	Patient(s) hospitalized for congestive heart failure
Treatment		
PPC-AD	Postpartum care rate	Women that received postpartum care (excluding bundled postpartum services)

Quality Measure Example

For readers who may not be familiar with quality measures, the following section provides an example. This example will use the Preventive Women's Health measure, breast cancer screening. Detailed measure specifications define the criteria for the Medicaid Adult Core Set measures and are available online.[83]

A key component of quality measures is identifying the people to whom the measure is genuinely applicable. For instance, the breast cancer screening measure only applies to women. Identifying the eligible population commonly employs Medicaid enrollment and/or claims data. This population is also commonly referred to as the denominator because it is used when calculating rates (i.e., percent of eligible population who satisfied the measure). The following are the criteria to identify the eligible population for the breast cancer screening measure.

- Age: Women ages 52 to 74 as of December 31 of the measurement year.
- Continuous Enrollment: October 1 two years prior to the measurement year through December 31 of the measurement year.
- Allowable Gap: No more than one gap in enrollment of up to 45 days for each full calendar year of continuous enrollment (i.e., the measurement year and the year prior to the measurement year). To determine continuous enrollment for a Medicaid enrollee for whom enrollment is verified monthly, the enrollee may not have more than a 1-month gap in coverage (i.e., an enrollee whose coverage lapses for 2 months [60 days] is not considered continuously enrolled). No gaps in enrollment are allowed from October 1 two years prior to the measurement year through December 31 two years prior to the measurement year.
- Anchor Date: December 31 of the measurement year.
- Benefit: Medical.

After identifying the eligible population, we next identify the people who satisfied the criteria for the measure. Identifying the people who satisfied the measure commonly employs claims data. This population is also commonly referred to as the numerator. In addition, when a person satisfies the criteria for the numerator, these instances are commonly referred to as a "yes" (e.g., yes, the person had a mammogram) whereas a person who did not satisfy the criteria are referred to as a "no" (e.g., no, the person did not have a mammogram). The following is the criterion to identify people who satisfied the breast cancer screening measure:

- One or more mammograms any time on or between October 1 two years prior to the measurement year and December 31 of the measurement year.

The measure specifications define the specific codes for each measure. For example, the breast cancer screening measure uses CPT codes (e.g., 77055 – Mammogram of one breast), HCPCS codes (e.g., G0202 – Screening mammography, producing direct digital image, bilateral, all views), and other codes (i.e., International Classification of Diseases, 9th Revision, Clinical Modification Procedure codes and Uniform Bill Revenue Codes) to identify a mammogram.

In addition, some measures have criteria to identify circumstances when a person should be excluded from a measure. For example, the breast cancer screening measure has an exclusion for bilateral mastectomy (i.e., surgical removal of all breast tissue to treat or prevent breast cancer).[84] That is, a woman who had this procedure could be excluded from the breast cancer screening measure. The measure

specifications define the criteria for exclusions including the applicable codes (e.g., CPT, HCPCS, etc.) to identify the exclusions.

Once the eligible population (i.e., denominator) and those satisfying the measure (i.e., numerator) are identified and any exclusions are omitted, this information can be used to derive a rate. This rate is commonly called an adherence rate. Specifically, the numerator is divided by the denominator to derive the rate. For example, if a given MCO had 100 members in their eligible population and 85 members who satisfied the measure, their adherence rate would be 85% (i.e., 85 / 100 = 85%).

B. Data Sources

Minnesota health care program data from nine Minnesota MCOs and FFS data were provided for this analysis, which encompassed Medicaid claims, enrollment, and provider data in addition to measures produced by the Johns Hopkins Adjusted Clinical Groups (ACG) system[79]; these data were provided by DHS and served as the foundation for many of the characteristics examined. The Lewin Group used this DHS data to derive new variables that assigned members into categories for the analysis. For example, DHS provided birthdate, which was used to calculate age and assign people to one of the age categories (i.e., under 45, 45 to 64, and 65+). Appendix E includes further detail about variable sources and definitions.

The Lewin Group derived the Medicaid Adult Core Set measure results; that is, this included identifying whether a person was eligible for and subsequently a "yes" or "no" for the select measures (see the Quality Measures section above for further description). Specifically, the measure results were produced by processing data provided by DHS (e.g., claims, enrollment, etc.) through the EBM Connect® software[81]. The data included calendar years 2011 through 2013. Data from years 2011 and 2012 were used mostly for the assessment of data quality (see Data Quality section below) while 2013 was used for the risk adjustment models. That is, while the 2013 measure results were used for the analyses, these measures commonly require a look-back period (e.g., two years prior to the measurement year) and therefore used data from the historical years for that purpose.

DHS uses the ACG system which provided multiple measures for this analysis, including expanded diagnosis cluster (EDC) measures[79]. Foremost, the ACGs provided the method of accounting for health risk with the categorical measure labeled resource utilization bands (RUBs). ACGs are categories (defined by morbidity, age, and gender) that group people with similar health or illness burden.[80] The RUBs combine people across ACG clinical categories into groups expected to have similar healthcare needs and resource utilization.[80] An indicator of frailty was also a measure produced by the ACGs. EDCs use diagnoses to categorize people with specific illnesses.[80] The EDC measures were used by the Lewin Group to derive the following clinical indicators: developmental disability, mental health, and substance abuse. The specific EDC values used to derive these indicators is available in Appendix E.

DHS' member enrollment data was the source for the sociodemographic characteristics, which included age, education, language, race and ethnicity, whether the member was enrolled in Medicaid due to a disability, and the county in which a member lived. Whether a member lived in a metropolitan county was derived by The Lewin Group by linking to the 2013 version of rural-urban continuum codes (RUCC), which is created by the U.S. Department of Agriculture.[85] Specifically, RUCC categories identified as Metro (i.e., values 1, 2, and 3) were categorized as metropolitan counties. Again, please see Appendix E for further detail about variable definitions.

Exclusion Criteria

There were two exclusion criteria applied to the data for all of the Medicaid Adult Core Set measures in this project. People who were enrolled in both Medicare and Medicaid (i.e., dual enrollees) were excluded. This information was obtained from Medicare Part A and Part B indicators from enrollment data provided by DHS. A person who was identified as having either Part A or Part B coverage was excluded. We did not have Medicare data for this analysis. Therefore, data were likely incomplete for these people, which could yield inaccurate results. The focus of this analysis was on adult measures; anyone under the age of 18 was excluded.

Data Quality

Several key steps were conducted to ensure the data was of sufficient quality to conduct the analysis. The first step was to check for data completeness. This was done by looking at claim counts, the number of unduplicated users, and the average claim per enrollee; each was examined by month of service and category of service. The second step was to check the validity of the data. This comprised of looking at member month distributions, member demographic distributions, and other enrollment statistics and comparing them year-over year. In addition, we examined the extent key variables had valid values, the extent provider and enrollment information linked to claims, and distributions of diagnoses. Based on these quality checks, we concluded the data quality was sufficient to proceed with the analysis.

C. Variables

We considered an assortment of characteristics for potential inclusion in risk adjustment models for each of the examined Medicaid Adult Core Set quality measures. In addition to characteristics such as clinical risk and age, which are commonly included in risk adjustment, a key interest was exploring other sociodemographic characteristics. We reviewed an earlier release of an NQF technical report related to this topic and examined a list of characteristics identified as potential sociodemographic characteristics for risk adjustment.[86] We subsequently identified characteristics present in DHS enrollment data with sufficient historical availability (i.e., 3 years). Later, DHS developed interests in potentially exploring additional characteristics. This encompassed other clinical characteristics (e.g., mental health) and utilization characteristics (e.g., emergency department use). Table 13 summarizes characteristics we considered, how those characteristics were categorized, and describes reasons some characteristics were not explored further. Additional detail about how characteristics were derived is available in Appendix E.

Table 13. Summary of variables considered for risk adjustment models.

Characteristic	Source	Categories	Notes
Clinical			
Developmental Disability	Derived from DHS EDC data	Yes or No	
Dialysis	Obtained from DHS ACG data	*N/A*	Not explored in models due to lack of variation (< 0.1% identified with dialysis)
Disability	Derived from DHS enrollment data	Yes or No	
Frailty	Derived from DHS ACG data	Yes or No	
Mental Health	Derived from DHS EDC data	Yes or No	

Characteristic	Source	Categories	Notes
Resource Utilization Bands (RUB)	Obtained from DHS ACG data	Non-user; Healthy; Low Morbidity; Moderate Morbidity; High Morbidity; Very High Morbidity	
Substance Abuse	Derived from DHS EDC data	Yes or No	
Sociodemographic			
Age	Derived from DHS enrollment data	Under 45; 45 to 64; 65+	
Education	Derived from DHS enrollment data	Unknown; Less than High School; High School Graduate; More than High School Graduate	
Gender	Derived from DHS enrollment data	Male or Female	
Immigration Status	Obtained from DHS enrollment data	*N/A*	Not explored in models due to concern regarding reliability of values
Language	Derived from DHS enrollment data	English or Non-English	
Metropolitan County	Derived from DHS enrollment data and USDA RUCCs[85]	Yes or No	
Race / Ethnicity	Derived from DHS enrollment data	White or Non-white	
Families and Children Medical Assistance Group	Derived from DHS enrollment data	*N/A*	Not included in models due to collinearity with the disability characteristic (please see Collinearity section below)
Utilization			Decision not to include utilization variables (please see Utilization Characteristics section below for rationale)
Emergency Department Visit	Derived from DHS ACG data	*N/A*	
Inpatient Hospitalization	Derived from DHS ACG data	*N/A*	
Major Procedure	Obtained from DHS ACG data	*N/A*	
Nursing	Obtained from DHS ACG data	*N/A*	
Outpatient Visit	Obtained from DHS ACG data	*N/A*	
Providers (Count)	Derived from DHS ACG data	*N/A*	
Saw a Specialist	Derived from DHS ACG data	*N/A*	
Saw a Generalist	Derived from DHS ACG data	*N/A*	

Variable Selection

Identifying the characteristics that were appropriate for each quality measure was an important next step in developing the risk adjustment methods to enhance use of the Medicaid Adult Core Set measures. That is, the next step was examining the candidate characteristics from this list and identifying those that should be included in the risk adjustment for each of the examined quality measures. More detail about the statistical variable selection methods is described in subsequent sections. Briefly, the variable selection process used both statistical criteria and non-statistical judgments. The statistical criteria effectively suggest which characteristics to include in a risk adjustment model for each of the quality measures. Even so, the statistical criteria need guidance and we also considered non-statistical judgments. Decisions were made to sometimes include a characteristic in a model even when it was not suggested by the statistical criteria. For example, including mental health in risk adjustment is of particular interest. The statistical criteria suggested inclusion of mental health in models for multiple quality measures, but not all. Given the demonstrated importance for some measures and the general importance of mental health, the decision was made to include this characteristic in models for other Medicaid Adult Core Set measures, when appropriate.

In addition, decisions were made not to include select characteristics when there was a direct relationship with the quality measure. The gender, mental health, and substance abuse characteristics were not examined for select quality measures. Gender was not examined for the three women's preventive health measures. The mental health indicator was not examined for the five mental health measures. Lastly, the substance abuse indicator was not examined for the two behavioral measures.

Utilization Characteristics

A number of utilization variables were characteristics considered for inclusion in the risk adjustment models; although, the decision was made not to include these utilization variables. This included whether someone had an emergency department visit, an inpatient hospitalization, a major procedure, a nursing service, an outpatient visit, the number of providers seen, whether someone had seen a specialist, and whether someone had seen a generalist.

The rationale for this decision was based on the underlying intent of risk adjusting these quality measures. Specifically, the aim of this risk adjustment was to account for factors an MCO cannot control thereby preventing improper impacts—negative or beneficial—to adherence rates. Natural examples are health and age. These are characteristics beyond the control of an MCO. That is, an MCO cannot control the age of members. While an MCO has some influence on health, the MCO cannot control the presence of preexisting conditions (e.g., heart disease, mental health, etc.). Utilization characteristics are not beyond the influence of an MCO. Moreover, utilization factors (e.g., hospitalization, number of providers seen, etc.) are inherently related to other factors already accounted for in the models (i.e., health risk and age).

Collinearity

We examined the candidate characteristics from the above list to assess the extent any of the variables were too closely related. Statistically, the analysis assessed the presence of collinearity. That is, collinearity is what statisticians call it when two characteristics are closely related (i.e., very high correlation). When this occurs and closely related characteristics are included in the same model, statistical problems can occur. For example, immigration status and language are characteristics that are potentially very closely related; people who immigrated to the U.S. might be more likely in general to not speak English. Therefore, these two characteristics could essentially be measuring something very

similar. While immigration status was excluded as a potential characteristic (see Table 13), this provides an example of variables that could be closely related.

Statistically, this can cause undesirable results such as unstable estimates, high standard errors, and large confidence intervals. This basically means that the results from a model might be misleading and we might have less trust in those results. Of course, we do not want this to happen and we conduct this analysis to prevent this from happening. This analysis examined all candidate variables that might be selected for the risk adjustment models. Therefore, if any were too closely related, we could identify the potential problem and select only one of the two variables for inclusion in the model to prevent such problems from arising.

Similar to the variable selection process described above, we use statistical criteria to guide the analysis. Specific statistical values provide suggestions when variables are too closely related. Specifically, the statistics we examined are called the variance inflation factor (VIF), condition index, and the proportion of variation. When a concern was suggested by these statistics, we examined additional values. Specifically, we examined frequency distribution cross-tabulations among the identified characteristics; this basically means we examined tables of counts and percentages. We explored whether the characteristics were closely related across the entire population used for this analysis. That is, we did not explore separate analyses for each group of people (i.e., the eligible population) for each quality measure. The rationale was that any relationships identified in the entire population would be similar in smaller groups of people from this population.

Ultimately, there was no evidence of collinearity among the predictor variables included in the final models. An accepted rule of thumb for interpreting the VIF is any value greater than 10 could indicate a collinearity concern.[87] The largest VIF value was 6.4. Similarly, an accepted rule of thumb for interpreting the condition index is a value greater than or equal to 30 could indicate a collinearity concern.[87] The largest condition index value was 6.9. Regarding the proportion of variation, multiple variables with values greater than 0.90 could indicate a collinearity concern.[87] There were no instances of multiple variables with values greater than 0.90. In summary, none of the collinearity results provided evidence of collinearity among variables included in final models.

Earlier analyses did identify a strong relationship between two variables. Members who were enrolled in Medicaid due to a disability and members enrolled in the Families and Children Medical Assistance group were highly correlated. Specifically, members enrolled in one category were not enrolled in the other (i.e., negative correlation). Therefore, we selected only one of these variables for further examination, enrollment due to a disability.

D. Risk Adjustment Models

The next step examined our candidate characteristics to assess the extent each was related with each of the Medicaid Adult Core Set measures. As described above in the Variables section, this process employed both statistical criteria and non-statistical decisions. The statistical criteria provide suggestions for which characteristics to include in the risk adjustment for each of the quality measures. The statistical criteria need guidance and are interpreted in combination with non-statistical judgements such as DHS' subject matter expertise.

The statistical method we used to guide the selection of characteristics to include in the risk adjustment for each quality measure is called logistic regression. We selected this method because it is appropriate for the type of data we examined. Specifically, logistic regression is an appropriate method for examining relationships when the outcome is dichotomous (i.e., only two potential options or outcomes). For example, considering the breast cancer screening measure, a person either was screened (i.e., a "yes"

outcome) or they were not (i.e., a "no" outcome). The "yes" and "no" outcomes were the only potential outcomes and therefore satisfies the criteria of having only two options (i.e., dichotomous). This was the case for all of the examined quality measures. The outcomes all had "yes" or "no" outcomes (e.g., a person had a service, etc.). In addition, the analysis was performed at what statisticians might call person-level. Each "yes" or "no" outcome is for an individual person. Again, a woman either was screened for breast cancer or she was not. We have information on the quality measures and all the candidate characteristics for individual people. This analysis then examined the information from those people to identify the relationships important for risk adjustment. The analysis of each quality measure included people who were appropriate for that measure (i.e., the eligible population). This included all people for whom we had data which encompassed both FFS and MCO members.

We used a method called stepwise variable selection to guide selection of the variables for risk adjustment. This method would run the analysis multiple times and use statistical criteria to make decisions each time. For example, the analysis would examine the candidate characteristics and select the variable with the strongest relationship—based on statistical criteria—with a given quality measure, and include that variable in the model. Next, the analysis would effectively repeat this process; the analysis would examine the remaining characteristics and select the next variable with the strongest relationship and add that variable to the model. A specific characteristic of stepwise variable selection is that variables can be added or removed at each step. For example, if when adding a new variable to the model a variable currently in the model no longer satisfies the statistical criteria, that previously entered variable would be removed. This process continues until none of the remaining variables have strong enough relationships to satisfy the criteria to enter the model, or all variables have already entered the model. Therefore, this process results in a set of characteristics that the statistics suggest should be included in the risk adjustment.

The entry criterion for a variable to enter the model was a 0.05 significance level. At each step, the variable with the largest test statistic entered the model if it satisfied this entry criterion. Similarly, for a predictor variable to remain in the model, it had to satisfy the retention criterion. The retention criterion was also a 0.05 significance level. In addition, we allowed the stepwise selection process to explore two multiplicative interactions. There were two multiplicative interactions of a priori interest: (1) age and health risk (i.e., resource utilization bands) and (2) race and health risk. While these were explored, neither was retained in final models.

Following the stepwise variable selection, we examined the models further with the aim of identifying the most parsimonious models. As mentioned previously, the statistical criteria require some guidance. Left unchecked, the statistics can include characteristics that, while satisfying these criteria, do not genuinely provide value. Therefore, we assessed the extent that variables were retained based on the above statistical significance criteria, but did not appear to add meaningfully to a model. To guide us in this step, we used another statistic that essentially tells us how well our model explains our outcome. Considering the breast cancer screening measure, this information would allow us to gauge how well the characteristics in a model explain whether women will be screened. When a characteristic was suggested by the above statistical criteria, but did not really improve a model's ability to explain the outcome, we considered removing that characteristic. Specifically, we selected a threshold based on what is called the R^2 statistic (pronounced R squared). This statistic can theoretically reach 100%, meaning that a model perfectly predicts the outcome. Our threshold for this value was 1/4% (i.e., 0.25). If adding a variable to the model did not improve our ability to explain the outcome by at least 1/4% (i.e., a very small amount) we considered excluding that variable.

Lastly, we finalized the models based on all the above statistical methods in addition to non-statistical considerations. The Minnesota Department of Human Services identified select characteristics for inclusion in final models based on their subject matter expertise and considerations appropriate for their

specific population(s). For example, the sociodemographic characteristics (e.g., race/ethnicity, language, etc.) were of particular interest to DHS. While the statistical criteria commonly suggested inclusion of these characteristics for risk adjustment for the quality measures, they were not always suggested. Given the general importance demonstrated by the inclusion in multiple models and DHS' interest, the decision was made to include these characteristics in additional models. Accordingly, we examined models that retained non-significant variables to ensure they were robust. That is, we examined the models to ensure including these characteristics did not cause any problems. We compared the models that included the non-significant variables to models that excluded those non-significant variables and assessed whether there were any abnormally large changes in select statistics. Specifically, we examined the stability of the model estimates (i.e., beta coefficients), the size and direction of the association, consistency of statistical significance and the risk adjusted rates. We found no evidence of any negative impact.

Risk Adjustment Model Example

The following section provides a simplified hypothetical example to illustrate the concept of how the risk adjustment methods are subsequently used to predict someone's outcome on a given quality measure. For this hypothetical example, we will continue using the Preventive Women's Health measure, breast cancer screening. Therefore, our model will want to predict if a woman will receive a mammogram. To keep this example simple, imagine we are examining the relationship of only one characteristic with this mammogram measure. That is, we want to know how a woman's age might help predict whether she receives a mammogram. Let's imagine that our statistical model indicates that women who are at least 65 years of age are less likely to receive a mammogram compared to younger women. Consider the following hypothetical results:

- Age (Hypothetical results for illustration)
    - 65 years of age and older: 50% chance of a mammogram
    - Under 65 Years of age: 75% chance of a mammogram

Therefore, when we apply this model to our eligible population for the breast cancer screening measure, women who are in the older age group would receive a probability score of 0.50 (i.e., 50% chance of a mammogram) whereas the younger age group would receive a score of 0.75. That is, for each person, we combine the information from the statistical model with a person's characteristics to derive a prediction.

While this was a simplified example only considering age, the statistical models that consider multiple characteristics are conceptually doing the same thing. That is, the models provide information about the relationships with the characteristics. We then combine that numerical information with a person's characteristics to derive a prediction. This prediction—commonly referred to as a score—is an important step that is then used in the next step of calculating risk adjusted rates.

E. Risk Adjusted Rates

In order to study the impact of risk adjusting the Medicaid Adult Core Set quality measures, we applied these methods to DHS' contracted MCOs. This allowed us to examine the extent accounting for the characteristics in the models altered the MCO's adherence rates on these quality measures.

The models were first used to derive a prediction for each person for a given quality measure. We call this prediction a score, which represents their probability of a "yes" or a "no" on a given quality measure. Again, this prediction is based on the variables included in the model and a person's specific characteristics for those variables. The score values range from 0 to 1, where 0 indicates a lower

probability of a "yes" and a 1 indicates a higher probability of a "yes." Therefore, this step takes the results of the models to give us a prediction for each person for a quality measure.

The next step takes these predictions for each person and we use that information to examine each MCO. That is, each person is a member in a specific MCO and we use the predictions from each person to make a prediction for all the people in an MCO. Specifically, the scores are summed for all members within each MCO for each quality measure. This value provides the estimate of what the statistics predict (i.e., what we call the *expected* rate) based on the model and an MCO's members' characteristics. This information is then combined with what really happened (i.e., what we call the *observed* rate). Ultimately, the calculation integrates the *observed* and *expected* rates with the overall population rate to derive the risk-adjusted rate for each MCO.

The following bullets summarize the key statistics related to the risk-adjusted adherence rates:

- Denominator: The unduplicated count of members satisfying the denominator criteria for a given quality measure for a group (e.g., MCO).
- Observed Count: The unduplicated count of members satisfying a given quality measure (i.e., members with a "yes") for a group (e.g., MCO).
- Expected Count: The count of members predicted to satisfy a given quality measure (i.e., sum of the probability scores) for a group (e.g., MCO).
- Observed Rate (O): The observed count divided by the denominator for a group (e.g., MCO).
- Expected Rate (E): The expected count divided by the denominator for a group (e.g., MCO).
- Population Rate: The observed count divided by the denominator for the entire population (i.e., regardless of MCO).
- Risk adjusted rate: The observed rate divided by the expected rate for a group, then multiplied by the population rate.
    - $(O/E) \times Population\ Rate$
- 95% Confidence Intervals:
    - $Risk\ adjusted\ rate\ \pm (1.96 \times Standard\ Error)$

While the statistics in the above bullets are fundamentally related for all the examined quality measures, several measures are slightly different. Specifically, the Chronic Hospitalization measures are based on the Prevention Quality Indicator (PQI) measures, which use member months as the denominator (i.e., instead of individuals). Member months is the number of months in a given year that a person was enrolled for benefits; this information was provided by DHS in their enrollment data. The PQI measures were calculated and displayed as a rate per 100,000 member months whereas the other measures' were calculated and displayed as a percent.

The above statistics aligned with methods published by the Agency for Healthcare Research and Quality.[88] Detailed equations can be found in Appendix D.

Risk Adjusted Rate Example

As described in the preceding risk adjustment models section, the statistical models are used to derive a prediction for each member that we call a score. This score represents a person's probability of a "yes" or a "no" on a given quality measure. To continue the breast cancer screening measure example from above, this score is the probability that a woman will receive a mammogram. The scores range between 0 and 1, where a 0 indicates a lower probability of a "yes" (e.g., having a mammogram) and a 1 indicates a higher

probability. These scores are then summed for all members within each MCO for each quality measure. Exhibit 1 contains a demonstration for two example groups or populations; for example, these groups could be considered two MCOs with different populations.

Exhibit 1. Example characteristics and scores for two populations.

Group	Person	Age Group	Score	Total
1	1	65+	0.50	2.0
1	2	65+	0.50	
1	3	65+	0.50	
1	4	65+	0.50	
2	1	< 65	0.75	3.0
2	2	< 65	0.75	
2	3	< 65	0.75	
2	4	< 65	0.75	

In this example, both groups have the same number of people (i.e., 4 people). However, we see that the ages of the people in the two groups are different. Group 1 has an older population (i.e., at least 65 years of age) whereas group 2 has a younger population (i.e., under 65 years of age). We continued the same hypothetical scores from these age groups from the previous section. That is, based on the hypothetical statistical model, the women in the older age group receive a score of 0.50 (i.e., 50% chance of a mammogram) whereas women in the younger age group receive a score of 0.75. We then sum the women's scores for each group.

- Total (i.e., sum of scores)
  - Group 1: (0.50 + 0.50 + 0.50 + 0.50) = 2.0
  - Group 2: (0.75 + 0.75 + 0.75 + 0.75) = 3.0

This value provides the estimate of how many members were expected to have a "yes" (e.g., a mammogram) for a given measure. That is, based on our statistical model and the age of the women in these groups, we expect 2 of the 4 women in group 1 (i.e., 50%) to receive a mammogram and 3 of the 4 women in group 2 (i.e., 75%). These concepts are further illustrated in Exhibit 2.

Exhibit 2. Example *expected* values for two populations.

Group	Denominator	Expected Count	Expected Rate
1	4	2	2 / 4 = 50%
2	4	3	3 / 4 = 75%

There are 4 people in each of these groups. This is the number that serves as the denominator for the *expected* rate. The sum of the scores provides the *expected* count, which serves as the numerator for the *expected* rate. Until now, we have only been considering what the statistical model tells us to expect. The next step in calculating the adjusted rate is examining what really happened (e.g., who received a mammogram). This is the information we actually *observe* (i.e., what really happened), and are accordingly labeled the *observed* count and rate. Exhibit 3 illustrates example *observed* results.

Exhibit 3. Example *observed* values for two populations.

Group	Denominator	Observed Count	Observed Rate
1	4	3	3 / 4 = 75%
2	4	2	2 / 4 = 50%

There are still 4 people in each of these groups. That is, the denominator is identical for both the *expected* and *observed* results. The count of women in each group who actually had a mammogram serves as the numerator for the *observed* rate. In this hypothetical example, 3 of the 4 women in group 1 received a mammogram (i.e., 75%) whereas 2 of the 4 women in group 2 (i.e., 50%) received a mammogram. Now we have the information about what we expect to happen (i.e., based on the model and the populations' characteristics) and what actually happened. Exhibit 4 illustrates how this information is brought together to calculate the risk adjusted rates.

Exhibit 4. Example risk adjusted rates for two populations.

Group	Observed Rate	Expected Rate	Population Rate	Risk Adjusted Rate
1	75%	50%	55%	0.75 / 0.50 * 0.55 = 82.5%
2	50%	75%	55%	0.50 / 0.75 * 0.55 = 36.7%

The risk adjusted rate is derived by first dividing a group's *observed* rate by the *expected* rate and then multiplying this value by the total population rate. In this hypothetical example, we are assuming that these two groups are part of a larger population and that this hypothetical population had a rate of 55%. For group 1, while we expected 50% of the women to receive a mammogram 75% did. That is, more women received a mammogram than we expected. Consequently, the risk adjusted rate for group 1 increased to 82.5% relative to the observed rate of 75%. For group 2, while we expected 75% of the women to receive a mammogram only 50% really did. That is, fewer women received a mammogram than we expected. Consequently, the risk adjusted rate for group 2 decreased to 36.7% relative to the observed rate of 50%.

This example demonstrates several key steps in the risk adjustment methods. It shows how scores are applied to individuals based on their characteristics (e.g., age). Next, we saw how those individuals and their scores relate to a group such as a MCO and predicting what we expect for that group. Lastly, we saw how we relate the statistical prediction (i.e., expected) to what actually happened (i.e., observed) to ultimately derive the risk adjusted rate.

Appendix B

This appendix provides information related to the phase one evaluation review where the Lewin Group risk adjusted select adult quality measures for a sample Medicaid population using ERGs and EBM Connect to explore particular features of certain measures that make the measures appropriate for risk adjustment. These results are submitted here as an example of how Lewin has previously risk adjusted quality measures. The methodology used is as follows:

Episode Risk Group (ERG) Methodology

1. Using medical encounter data, pharmacy encounter data, and member enrollment data, health care services for each member are assigned to unique Episodes of Care, called Episode Treatment Groups (ETG). ETGs are composed primarily of anchor records, which are records that demonstrate that the clinician has determined which further services are required to treat a medical condition (i.e., a claim for services related evaluation of a member's condition, a claim for surgical procedures, or a claim submitted for emergency room services or treatment facilities) Other records, such as those for x-rays, lab tests, and pharmaceuticals, are grouped to a clinically appropriate anchor record, creating clusters. ETGs are able to prioritize related medical conditions and services within the group, shifting focus to the condition that is most representative of the unique mix of services in the ETG.
2. ETGs are then further categorized into one of 167 ERGs. ERGs are essentially markers of member risk and aggregate ETGs of similar clinical and risk attributes. The severity levels in each ETG are analyzed to determine the pattern of utilization suggests that different levels of risk are present. Members may be assigned to zero, one, or several ERGs.
3. Each member's ERG clinical and demographic risk profile is developed based on age, sex, and mix of ERGs.
4. Finally, a retroactive risk score is computed by summing the predetermined weights attached to each ERG and demographic characteristic. The risk weights for the ERG model determined using multiple linear regression and enrollment, medical, and pharmacy claims data for a large managed care population.

Evidence-Based Medicine (EBM) Connect Methodology

EBM Connect links national standard quality measures and measures that have undergone extensive clinical research and review to provide information for health care quality measurement. When run in conjunction with ETG, EBM Connect can support direct comparison of quality and cost outcomes for a plethora of clinical conditions. It uses measure standard specifications from national health care quality and performance measures from the American Medical Association Physician Consortium for Performance Improvement (PCPI), CMS Physician Quality Reporting Initiative (PQRI), and NCQA HEDIS, with priority given to National Quality Forum (NCQ)-endorsed measures. The most current version of EBM Connect contains over 640 measures across 74 clinical conditions. Clinical measures are organized around five conceptual categories: disease management, medication adherence, national standards, patient safety, and care pattern.

Statistical significance

A logistic regression model of acuity level on adherence was created, and the statistical significance of the association between member acuity and adherence was assessed using the p-value. If the p-value for any acuity level in the measure was not statistically significant ($p>0.05$ for this analysis), the impact of risk adjustment was deemed not significant overall. If the adherence rates did not increase with the acuity level, the impact was also determined to be less significant.

Findings

Group 1: Table 14 to Table 18 display the results of risk adjusting a general set of adult quality measures using a sample Medicaid population. In Table 14 and Table 15 the level of adherence is relatively consistent across all acuity levels, suggesting that risk adjustment on the basis of acuity may not be particularly useful. Conversely, in Table 16, Table 17, and Table 18, adherence levels vary dramatically with acuity level, suggesting that that risk adjustment for acuity would be beneficial. However, the small sample size of medium and low acuity members must be taken into consideration when interpreting the results. In subsequent risk adjustment analyses, Lewin will recalculate the acuity levels based on population characteristics and will attempt to achieve a normal distribution.

Table 14. Risk adjustment for EBM Connect measure, patient(s) 21 - 24 years of age that had a chlamydia screening test in last 12 reported months.

Acuity level	Adherent members	Non-adherent members	Total members included	Percent adherent	Statistical Significance
Very high	1,163	685	1,848	62.93%	<.0001
High	1,568	1,183	2,751	57.00%	<.0001
Medium	985	900	1,885	52.26%	<.0001
Low	304	429	733	41.47%	N/A
Totals	4,020	3,197	7,217	55.70%	

Table 15. Risk adjustment for EBM Connect measure, patient(s) that had a cervical cancer screening test in last 36 reported months.

Acuity level	Adherent members	Non-adherent members	Total members included	Percent adherent	Statistical Significance
Very high	4,587	3,226	7,813	58.71%	<.0001
High	4,385	2,488	6,873	63.80%	<.0001
Medium	2,789	1,826	4,615	60.43%	<.0001
Low	982	1,205	2,187	44.90%	N/A
Totals	12,743	8,745	21,488	59.30%	

Table 16. Risk adjustment for EBM Connect measure, patient(s) 42 - 69 years of age that had a screening mammogram in last 24 reported months.

Acuity level	Adherent members	Non-adherent members	Total members included	Percent adherent	Statistical Significance
Very high	2,219	2,381	4,600	48.24%	<.0001
High	1,061	1,522	2,583	41.08%	<.0001
Medium	377	816	1,193	31.60%	<.0001
Low	69	741	810	8.52%	N/A
Totals	3,726	5,460	9,186	40.56%	

Table 17. Risk adjustment for EBM Connect measure, patient(s) 18 - 75 years of age that had an HbA1c test in last 12 reported months.

Acuity level	Adherent members	Non-adherent members	Total members included	Percent adherent	Statistical Significance
Very high	3,101	911	4,012	77.29%	<.0001
High	1,178	499	1,677	70.24%	<.0001
Medium	276	207	483	57.14%	<.0001
Low	2	55	57	3.51%	N/A
Totals	4,557	1,672	6,229	73.16%	

Table 18. Risk adjustment for EBM Connect measure, patient(s) 18 - 75 years of age with LDL cholesterol in last 12 months.

Acuity level	Adherent members	Non-adherent members	Total members included	Percent adherent	Statistical Significance
Very high	2,478	1,534	4,012	61.77%	<.0001
High	925	752	1,677	55.16%	<.0001
Medium	193	290	483	39.96%	N/A
Low	0	57	57	0.00%	N/A
Totals	3,596	2,633	6,229	57.73%	

Group 2: Table 19 and Table 20 show results for medication adherence measures. The member distribution for both measures falls into a limited spectrum of one or two acuity levels with nearly uniform adherence levels across the major acuity levels, indicating that risk adjusting these measures may not be necessary.

Table 19. Risk adjustment for EBM Connect measure, patient(s) with schizophrenia who remained on antipsychotic medication for at least 80% of their treatment period.

Acuity level	Adherent members	Non-adherent members	Total members included	Percent adherent	Statistical Significance
Very high	611	249	860	71.05%	0.8293
High	321	142	463	69.33%	0.9737
Medium	32	14	46	69.57%	N/A
Totals	964	405	1,369	70.42%	

Table 20. Risk adjustment for EBM Connect measure, adult patients persistently taking digoxin containing medication who received a serum potassium test AND either a serum creatinine or a blood urea nitrogen test within the last 12 reported months (HEDIS criteria).

Acuity level	Adherent members	Non-adherent members	Total members included	Percent adherent	Statistical Significance
Very high	93	2	95	97.90%	<.0001
High	9	4	13	69.23%	N/A
Medium	1	0	1	100.00%	N/A
Totals	103	6	109	94.50%	

Group 3: Table 21 shows the results of the analysis for two follow-up measures. In these cases, members in high or very high acuity levels are clearly more adherent than members in medium and low acuity groups, suggesting that risk adjustment should be applied.

Table 21. Risk adjustment for EBM Connect measure, patient(s) hospitalized for mental illness that had a follow-up encounter with a mental health practitioner within 30 days after discharge.

Acuity level	Adherent members	Non-adherent members	Total members included	Percent adherent	Statistical Significance
Very high	965	643	1,608	60.01%	<.0001
High	1,000	916	1,916	52.19%	0.0042
Medium	391	451	842	46.44%	0.0821
Low	36	61	97	37.11%	N/A
Totals	2,392	2,071	4,463	53.60%	

f characteristics included in final models, by MCO.

	MCO 9	MCO 8	MCO 7	MCO 6	MCO 5	MCO 4	MCO 3	MCO 2	MCO 1	FFS
	138 (7.9%)	85 (0.9%)	383 (0.7%)	2,031 (1.6%)	1,497 (1.9%)	144 (1.1%)	20 (0.5%)	602 (0.7%)	180 (1.5%)	5,748 (3.6%)
	1,611 (92.1%)	9,783 (99.1%)	56,062 (99.3%)	127,171 (98.4%)	77,661 (98.1%)	12,667 (98.9%)	3,910 (99.5%)	79,770 (99.3%)	11,705 (98.5%)	156,042 (96.4%)
	1,627 (93%)	67 (0.7%)	226 (0.4%)	9,501 (7.4%)	8,219 (10.4%)	616 (4.8%)	22 (0.6%)	214 (0.3%)	596 (5%)	30,222 (18.7%)
	122 (7%)	9,801 (99.3%)	56,219 (99.6%)	119,701 (92.6%)	70,939 (89.6%)	12,195 (95.2%)	3,908 (99.4%)	80,158 (99.7%)	11,289 (95%)	131,568 (81.3%)
	378 (21.6%)	895 (9.1%)	3,123 (5.5%)	8,679 (6.7%)	5,862 (7.4%)	870 (6.8%)	205 (5.2%)	4,444 (5.5%)	723 (6.1%)	10,118 (6.3%)
	1,371 (78.4%)	8,973 (90.9%)	53,322 (94.5%)	120,523 (93.3%)	73,296 (92.6%)	11,941 (93.2%)	3,725 (94.8%)	75,928 (94.5%)	11,162 (93.9%)	151,672 (93.7%)
	994 (56.8%)	3,038 (30.8%)	12,335 (21.9%)	32,736 (25.3%)	22,059 (27.9%)	3,305 (25.8%)	1,133 (28.8%)	17,757 (22.1%)	2,995 (25.2%)	37,946 (23.5%)
	755 (43.2%)	6,830 (69.2%)	44,110 (78.1%)	96,466 (74.7%)	57,099 (72.1%)	9,506 (74.2%)	2,797 (71.2%)	62,615 (77.9%)	8,890 (74.8%)	123,844 (76.5%)
s (RUB)										
	42 (2.4%)	2,450 (24.8%)	6,238 (11.1%)	13,698 (10.6%)	7,894 (10%)	1,164 (9.1%)	381 (9.7%)	7,561 (9.4%)	1,135 (9.5%)	29,541 (18.4%)
	114 (6.5%)	1,406 (14.2%)	8,912 (15.8%)	18,015 (13.9%)	10,568 (13.4%)	1,750 (13.7%)	535 (13.6%)	11,934 (14.8%)	1,611 (13.6%)	40,408 (25.1%)

Characteristics	MCO 9	MCO 8	MCO 7	MCO 6	MCO 5	MCO 4	MCO 3	MCO 2	MCO 1	FFS
Moderate	691 (39.5%)	3,372 (34.2%)	23,648 (41.9%)	52,747 (40.8%)	32,808 (41.4%)	5,295 (41.3%)	1,717 (43.7%)	34,796 (43.3%)	4,937 (41.5%)	44,896 (27.9%)
High	443 (25.3%)	1,072 (10.9%)	8,010 (14.2%)	22,935 (17.8%)	13,923 (17.6%)	2,404 (18.8%)	714 (18.2%)	12,327 (15.3%)	2,226 (18.7%)	20,540 (12.8%)
Very High	348 (19.9%)	499 (5.1%)	1,828 (3.2%)	6,131 (4.7%)	4,560 (5.8%)	543 (4.2%)	144 (3.7%)	2,700 (3.4%)	407 (3.4%)	8,428 (5.2%)
Substance Abuse										
Yes	876 (50.1%)	2,917 (29.6%)	12,952 (22.9%)	32,824 (25.4%)	22,086 (27.9%)	3,787 (29.6%)	1,229 (31.3%)	19,128 (23.8%)	3,429 (28.9%)	31,201 (19.3%)
No	873 (49.9%)	6,951 (70.4%)	43,493 (77.1%)	96,378 (74.6%)	57,072 (72.1%)	9,024 (70.4%)	2,701 (68.7%)	61,244 (76.2%)	8,456 (71.1%)	130,589 (80.7%)
Sociodemographic										
Age										
< 45	732 (41.9%)	6,700 (67.9%)	41,241 (73.1%)	93,826 (72.6%)	55,275 (69.8%)	9,699 (75.7%)	2,784 (70.8%)	56,352 (70.1%)	9,087 (76.5%)	129,801 (80.2%)
45 – 64	910 (52%)	3,165 (32.1%)	14,726 (26.1%)	33,826 (26.2%)	23,146 (29.2%)	3,085 (24.1%)	1,136 (28.9%)	23,541 (29.3%)	2,764 (23.3%)	30,586 (18.9%)
65+	107 (6.1%)	3 (0%)	478 (0.8%)	1,550 (1.2%)	737 (0.9%)	27 (0.2%)	10 (0.3%)	479 (0.6%)	34 (0.3%)	1,403 (0.9%)
Education										
Unknown	5 (0.3%)	114 (1.2%)	900 (1.6%)	1,726 (1.3%)	936 (1.2%)	206 (1.6%)	34 (0.9%)	1,601 (2%)	174 (1.5%)	3,117 (1.9%)
Less than High School	851 (48.7%)	3,176 (32.2%)	19,343 (34.3%)	49,432 (38.3%)	26,840 (33.9%)	3,865 (30.2%)	859 (21.9%)	24,885 (31%)	3,564 (30%)	73,150 (45.2%)
High School Graduate	775 (44.3%)	5,435 (55.1%)	28,921 (51.2%)	63,633 (49.3%)	42,258 (53.4%)	7,297 (57%)	2,608 (66.4%)	43,552 (54.2%)	6,943 (58.4%)	69,744 (43.1%)
More Than High School Graduate	118 (6.7%)	1,143 (11.6%)	7,281 (12.9%)	14,411 (11.2%)	9,124 (11.5%)	1,443 (11.3%)	429 (10.9%)	10,334 (12.9%)	1,204 (10.1%)	15,779 (9.8%)
Gender										
Male	956 (54.7%)	7,440 (75.4%)	24,388 (43.2%)	54,523 (42.2%)	32,572 (41.1%)	5,256 (41%)	1,745 (44.4%)	32,988 (41%)	4,774 (40.2%)	57,168 (35.3%)
Female	793 (45.3%)	2,428 (24.6%)	32,057 (56.8%)	74,679 (57.8%)	46,586 (58.9%)	7,555 (59%)	2,185 (55.6%)	47,384 (59%)	7,111 (59.8%)	104,622 (64.7%)

Characteristics	MCO 9	MCO 8	MCO 7	MCO 6	MCO 5	MCO 4	MCO 3	MCO 2	MCO 1	FFS
Language										
English	1,475 (84.3%)	9,101 (92.2%)	48,513 (85.9%)	101,882 (78.9%)	70,629 (89.2%)	12,389 (96.7%)	3,792 (96.5%)	73,333 (91.2%)	11,364 (95.6%)	139,889 (86.5%)
Non-English	274 (15.7%)	767 (7.8%)	7,932 (14.1%)	27,320 (21.1%)	8,529 (10.8%)	422 (3.3%)	138 (3.5%)	7,039 (8.8%)	521 (4.4%)	21,901 (13.5%)
Metropolitan County										
Yes	1,725 (98.6%)	9,654 (97.8%)	55,463 (98.3%)	111,101 (86%)	68,652 (86.7%)	242 (1.9%)	70 (1.8%)	41,531 (51.7%)	1,747 (14.7%)	118,042 (73%)
No	2 (1.4%)	214 (2.2%)	982 (1.7%)	18,101 (14%)	10,506 (13.3%)	12,569 (98.1%)	3,860 (98.2%)	38,841 (48.3%)	10,138 (85.3%)	43,748 (27%)
Race / Ethnicity										
White	390 (22.3%)	2,850 (28.9%)	27,607 (48.9%)	66,177 (51.2%)	45,779 (57.8%)	10,122 (79%)	3,444 (87.6%)	64,061 (79.7%)	10,120 (85.1%)	98,049 (60.6%)
Non-white	1,359 (77.7%)	7,018 (71.1%)	28,838 (51.1%)	63,025 (48.8%)	33,379 (42.2%)	2,689 (21%)	486 (12.4%)	16,311 (20.3%)	1,765 (14.9%)	63,741 (39.4%)

B. Risk Adjustment Models

Interpretation

To provide an example how these values can be interpreted, we will again focus on the breast cancer screening measure. The following provides descriptions of the values and examples from the breast cancer screening measure in Table 23.

- Parameter Estimates: These values provide information about the direction and size of the association.
    - Direction of association: Statistically, we look at the direction of the association to determine if a characteristic had what we call a positive or negative association. For the breast cancer screening measure, this tells us whether people with a given characteristic were more likely to have had a mammogram (i.e., positive association) or less likely (i.e., negative association). Values less than zero (i.e., values preceded by a negative symbol) represent a negative association. Values greater than zero represent a positive association.
        - Examples:
            - Developmental Disability: The column labeled "Level" indicates "No vs. Yes", which tells us more about how to interpret these values. This means we are comparing women without a developmental disability (i.e., a "No" for this characteristic) to women with a developmental disability (i.e., a "Yes" for this characteristic). We see that the estimate is negative (i.e., -0.08), which means women without a developmental disability were less likely to have received a mammogram. Even so, this characteristic did not have a statistically significant influence on this measure (discussed below). You will note this is the exact interpretation provided by the odds ratio (discussed below).
            - Age: The level column indicates we are comparing younger women (i.e., ages 45-64) to older women (i.e., ages 65+). We see this value is positive (i.e., 0.50), which means younger women were more likely to have received a mammogram. Age had a statistically significant impact.
    - Size of Association: By "size", we mean that some characteristics have a bigger impact than others. In short, the bigger the value, the bigger the impact. We must also consider the direction of association (e.g., bigger negative or bigger positive association).
        - Examples:
            - Above we saw that not having a developmental disability had a negative association (i.e. -0.08) with receiving a mammogram. We see that language also had a negative association (i.e., -0.25); non-English speakers were less likely to have had a mammogram. Language had a larger negative value compared to developmental disability (i.e., -0.25 is a larger negative value than -0.08); therefore, language had a larger impact on whether women received a mammogram compared to a developmental disability.

- Above we saw that younger age had a positive association (i.e., 0.50) with receiving a mammogram. We see that frailty also had a positive association (i.e., 0.09); women who were not frail were more likely to have had a mammogram. Age had a larger value compared to frailty (i.e., 0.50 is larger than 0.09); therefore, age had a larger impact on whether women received a mammogram.

- Odds Ratio (OR): These values are closely related to the parameter estimates (and are actually calculated directly from the parameter estimates). Odds ratios provide the same type of information regarding the direction and size of the association.

    - Direction of association: Similar to the parameter estimate, the odds ratio tells whether people with a given characteristic had a positive association (e.g., more likely to have had a mammogram) or a negative association (e.g., less likely to have had a mammogram). Values less than one (i.e., 1.0) represent a negative association. Values greater than one represent a positive association.

        - Examples:

            - Developmental Disability: Again, the level indicates we are comparing women without a developmental disability (i.e., "No" on this characteristic) to women with a disability (i.e., "Yes" on this characteristic). We see that the odds ratio is less than one (i.e., OR = 0.92), which means women without a developmental disability were less likely to have received a mammogram. You will note this is the exact interpretation provided by the parameter estimate (discussed above).

            - Age: The level indicates we are comparing younger women (i.e., ages 45-64) to older women (i.e., ages 65+). We see the odds ratio is greater than one (i.e., OR = 1.65), which means younger women were more likely to have received a mammogram.

    - Size of association: Similar to the parameter estimate, the size of the odds ratio indicates whether a characteristic had a bigger impact. For characteristics with a negative association (e.g., less likely to have a mammogram), a smaller number (i.e., closer to zero) indicates a larger negative impact. For characteristics with a positive association (e.g., more likely to have a mammogram), a larger number (i.e., greater and further from 1.0) indicates a larger positive impact.

        - Examples:

            - Again, we saw that having a developmental disability had a negative association (i.e., OR = 0.92) with receiving a mammogram. We see that language also had a negative association (i.e., OR = 0.78). Language had a smaller number compared to developmental disability (i.e., 0.78 is smaller than 0.92); therefore, language had a larger impact on whether women received a mammogram compared to developmental disability.

            - Again, we saw that younger age had a positive association (i.e., OR = 1.65) with receiving a mammogram. We see that frailty also had a positive association (i.e., OR = 1.10); women who were not frail were

more likely to have had a mammogram. Age had a larger value compared to frailty (i.e., 1.65 is larger than 1.10); therefore, age had a larger impact on whether women received a mammogram.

- P-values: these values represent statistical significance. In short, statistical significance provides a gauge of how much we trust the results. When a result is statistically significant, we tend to have more confidence the result is genuine. When a result is not statistically significant, this could mean the result is possibly just coincidental or what we call random chance. Values less than 0.05 are commonly considered statistically significant. Therefore, even smaller values (e.g., <0.0001) are also—some say more—statistically significant. We presented two separate p-values.
    - Category: This p-value represents whether a given category within a given characteristic was statistically significant.
        - Examples:
            - Resource Utilization Bands:
                - We see the p-value for women with high morbidity was not statistically significant (i.e., 0.1709 is greater than 0.05). The level for this characteristic indicates that each of these health risk categories was compared to women with very high morbidity. Therefore, women with high morbidity were not significantly different from women with very high morbidity regarding whether they would receive a mammogram.
                - We also see the p-value for women with moderate morbidity was statistically significant (i.e., <0.0001 is much less than 0.05). Therefore, women with moderate morbidity were significantly less likely to have a mammogram compared to very high morbidity women.
    - Variable: This p-value represents whether the entire variable was statistically significant in the model.
        - Examples:
            - Developmental Disability: we see the p-value for developmental disability was not statistically significant (i.e., 0.4772 was greater than 0.05). Therefore, this characteristic did not have a statistically significant influence on whether women received a mammogram.
            - Resource Utilization Bands: we see the p-value for this health risk measure was statistically significant (i.e., <0.0001 is much less than 0.05). Therefore, health risk had a statistically significant influence on whether women received a mammogram.

Table 23. Risk adjustment model results for the breast cancer screening measure (BCS-AD).

Parameter	Level	Estimate	Odds Ratio	Category p-value	Variable p-value
Intercept		-0.01	0.99	0.9580	
Developmental Disability	No vs. Yes	-0.08	0.92	0.4772	0.4772
Disability	No vs. Yes	0.47	1.60	<.0001	<.0001
Frailty	No vs. Yes	0.09	1.10	0.0376	0.0376
Mental Health	No vs. Yes	0.00	1.00	0.8911	0.8911
Resource Utilization Bands (RUB)	Non-user vs. Very High	-3.32	0.04	<.0001	<.0001
Resource Utilization Bands (RUB)	Healthy User vs. Very High	-1.41	0.25	<.0001	
Resource Utilization Bands (RUB)	Low Morbidity vs. Very High	-0.99	0.37	<.0001	
Resource Utilization Bands (RUB)	Moderate Morbidity vs. Very High	-0.25	0.78	<.0001	
Resource Utilization Bands (RUB)	High Morbidity vs. Very High	0.07	1.07	0.1709	
Substance Abuse	No vs. Yes	0.04	1.04	0.3128	0.3128
Age	45-64 vs. 65+	0.50	1.65	<.0001	<.0001
Education	Unknown vs. More than High School	0.24	1.27	0.0846	0.0165
Education	Less than High School vs. More than High School	-0.05	0.95	0.3465	
Education	High School Graduate vs. More than High School	0.05	1.05	0.2928	
Language	Non-English vs. English	-0.25	0.78	<.0001	<.0001
Metropolitan County	No vs. Yes	-0.04	0.96	0.2661	0.2661
Race / Ethnicity	Non-white vs. White	-0.04	0.96	0.2332	0.2332

Table 24. Risk adjustment model results for the cervical cancer screening measure (CCS-AD).

Parameter	Level	Estimate	Odds Ratio	Category p-value	Variable p-value
Intercept		0.18	1.20	0.0013	
Developmental Disability	No vs. Yes	0.48	1.62	<.0001	<.0001
Disability	No vs. Yes	0.73	2.07	<.0001	<.0001
Mental Health	No vs. Yes	0.03	1.03	0.1043	0.1043
Resource Utilization Bands (RUB)	Non-user vs. Very High	-1.92	0.15	<.0001	<.0001
Resource Utilization Bands (RUB)	Healthy User vs. Very High	-0.81	0.45	<.0001	
Resource Utilization Bands (RUB)	Low Morbidity vs. Very High	-0.38	0.68	<.0001	
Resource Utilization Bands (RUB)	Moderate Morbidity vs. Very High	-0.02	0.98	0.5159	
Resource Utilization Bands (RUB)	High Morbidity vs. Very High	0.32	1.38	<.0001	
Substance Abuse	No vs. Yes	-0.04	0.96	0.0253	0.0253
Age	45-64 vs. Under 45	-0.83	0.43	<.0001	<.0001
Education	Unknown vs. More than High School	-0.06	0.95	0.5269	<.0001
Education	Less than High School vs. More than High School	-0.19	0.83	<.0001	
Education	High School Graduate vs. More than High School	-0.09	0.91	0.0001	
Language	Non-English vs. English	-0.10	0.91	0.0002	0.0002
Metropolitan County	No vs. Yes	-0.23	0.79	<.0001	<.0001
Race / Ethnicity	Non-white vs. White	0.08	1.08	<.0001	<.0001

Table 25. Risk adjustment model results for the chlamydia screening in women (CHL-AD).

Parameter	Level	Estimate	Odds Ratio	Category p-value	Variable p-value
Intercept		-0.61	0.54	0.0002	
Developmental Disability	No vs. Yes	0.65	1.92	<.0001	<.0001
Disability	No vs. Yes	0.81	2.25	<.0001	<.0001
Mental Health	No vs. Yes	-0.25	0.78	<.0001	<.0001
Resource Utilization Bands (RUB)	Non-user vs. Very High	-10.71	0.00	0.9275	<.0001
Resource Utilization Bands (RUB)	Healthy User vs. Very High	-0.82	0.44	<.0001	
Resource Utilization Bands (RUB)	Low Morbidity vs. Very High	-0.25	0.78	0.0857	
Resource Utilization Bands (RUB)	Moderate Morbidity vs. Very High	-0.10	0.90	0.4547	
Resource Utilization Bands (RUB)	High Morbidity vs. Very High	-0.08	0.92	0.5521	
Substance Abuse	No vs. Yes	-0.30	0.74	<.0001	<.0001
Education	Unknown vs. More than High School	-0.05	0.95	0.8019	0.0562
Education	Less than High School vs. More than High School	0.15	1.16	0.0173	
Education	High School Graduate vs. More than High School	0.16	1.17	0.0128	
Language	Non-English vs. English	-0.31	0.73	<.0001	<.0001
Metropolitan County	No vs. Yes	-0.49	0.61	<.0001	<.0001
Race / Ethnicity	Non-white vs. White	0.59	1.80	<.0001	<.0001

Table 26. Risk adjustment model results for the Annual monitoring for enrollees on angiotensin converting enzyme (ACE) inhibitors or angiotensin receptor blockers (ARB) measure (MPM-AD-R1).

Parameter	Level	Estimate	Odds Ratio	Category p-value	Variable p-value
Intercept		3.46	31.77	<.0001	
Developmental Disability	No vs. Yes	0.05	1.05	0.7279	0.7279
Disability	No vs. Yes	0.13	1.13	0.0147	0.0147
Frailty	No vs. Yes	-0.48	0.62	<.0001	<.0001
Mental Health	No vs. Yes	0.01	1.02	0.7818	0.7818
Resource Utilization Bands (RUB)	Non-user vs. Very High	-13.11	0.00	0.9144	<.0001
Resource Utilization Bands (RUB)	Healthy User vs. Very High	-3.44	0.03	<.0001	
Resource Utilization Bands (RUB)	Low Morbidity vs. Very High	-2.67	0.07	<.0001	
Resource Utilization Bands (RUB)	Moderate Morbidity vs. Very High	-1.85	0.16	<.0001	
Resource Utilization Bands (RUB)	High Morbidity vs. Very High	-1.17	0.31	<.0001	
Substance Abuse	No vs. Yes	-0.05	0.95	0.3326	0.3326
Age	45-64 vs. Under 45	0.37	1.45	<.0001	<.0001
Age	65+ vs. Under 45	0.73	2.07	<.0001	
Gender	Female vs. Male	0.08	1.08	0.0738	0.0738
Education	Unknown vs. More than High School	0.42	1.52	0.0254	0.1597
Education	Less than High School vs. More than High School	0.01	1.01	0.8508	
Education	High School Graduate vs. More than High School	0.01	1.01	0.8261	
Language	Non-English vs. English	0.22	1.25	0.0019	0.0019
Metropolitan County	No vs. Yes	-0.11	0.90	0.0226	0.0226
Race / Ethnicity	Non-white vs. White	-0.03	0.97	0.5681	0.5681

Table 27. Risk adjustment model results for the Annual monitoring for enrollees on diuretics measure (MPM-AD-R3).

Parameter	Level	Estimate	Odds Ratio	Category p-value	Variable p-value
Intercept		4.00	54.56	<.0001	
Developmental Disability	No vs. Yes	-0.29	0.75	0.1349	0.1349
Disability	No vs. Yes	0.12	1.12	0.0533	0.0533
Frailty	No vs. Yes	-0.46	0.63	<.0001	<.0001
Mental Health	No vs. Yes	0.10	1.10	0.1139	0.1139
Resource Utilization Bands (RUB)	Healthy User vs. Very High	-3.72	0.02	<.0001	<.0001
Resource Utilization Bands (RUB)	Low Morbidity vs. Very High	-2.91	0.05	<.0001	
Resource Utilization Bands (RUB)	Moderate Morbidity vs. Very High	-2.06	0.13	<.0001	
Resource Utilization Bands (RUB)	High Morbidity vs. Very High	-1.31	0.27	<.0001	
Substance Abuse	No vs. Yes	-0.07	0.93	0.2069	0.2069
Age	45-64 vs. Under 45	0.39	1.48	<.0001	<.0001
Age	65+ vs. Under 45	0.74	2.10	<.0001	
Gender	Female vs. Male	-0.08	0.92	0.0963	0.0963
Education	Unknown vs. More than High School	0.59	1.81	0.0111	0.0328
Education	Less than High School vs. More than High School	0.11	1.11	0.1800	
Education	High School Graduate vs. More than High School	0.02	1.02	0.7894	
Language	Non-English vs. English	0.17	1.18	0.0534	0.0534
Metropolitan County	No vs. Yes	-0.14	0.87	0.0121	0.0121
Race / Ethnicity	Non-white vs. White	-0.05	0.95	0.3365	0.3365

Table 28. Risk adjustment model results for the Annual monitoring for enrollees on anticonvulsants measure (MPM-AD-R4).

Parameter	Level	Estimate	Odds Ratio	Category p-value	Variable p-value
Intercept		2.44	11.43	<.0001	
Developmental Disability	No vs. Yes	-0.36	0.70	0.0002	0.0002
Disability	No vs. Yes	-0.04	0.96	0.6417	0.6417
Frailty	No vs. Yes	-0.24	0.79	0.0101	0.0101
Mental Health	No vs. Yes	-0.12	0.88	0.1327	0.1327
Resource Utilization Bands (RUB)	Healthy User vs. Very High	-1.37	0.25	<.0001	<.0001
Resource Utilization Bands (RUB)	Low Morbidity vs. Very High	-1.58	0.21	<.0001	
Resource Utilization Bands (RUB)	Moderate Morbidity vs. Very High	-1.02	0.36	<.0001	
Resource Utilization Bands (RUB)	High Morbidity vs. Very High	-0.69	0.50	<.0001	
Substance Abuse	No vs. Yes	-0.21	0.81	0.0077	0.0077
Age	45-64 vs. Under 45	-0.03	0.97	0.6959	0.1000
Age	65+ vs. Under 45	1.13	3.10	0.0373	
Gender	Female vs. Male	-0.13	0.88	0.0630	0.0630
Education	Unknown vs. More than High School	-0.15	0.86	0.6302	0.9511
Education	Less than High School vs. More than High School	-0.06	0.94	0.6408	
Education	High School Graduate vs. More than High School	-0.04	0.96	0.7258	
Language	Non-English vs. English	0.15	1.17	0.2829	0.2829
Metropolitan County	No vs. Yes	-0.17	0.85	0.0374	0.0374
Race / Ethnicity	Non-white vs. White	0.01	1.01	0.9202	0.9202

Table 29. Risk adjustment model results for the Comprehensive diabetes care Hemoglobin A1c testing measure (HA1C-AD).

Parameter	Level	Estimate	Odds Ratio	Category p-value	Variable p-value
Intercept		1.46	4.32	<.0001	
Developmental Disability	No vs. Yes	-0.12	0.89	0.2845	0.2845
Disability	No vs. Yes	0.25	1.29	<.0001	<.0001
Frailty	No vs. Yes	-0.07	0.94	0.2134	0.2134
Mental Health	No vs. Yes	0.20	1.23	<.0001	<.0001
Resource Utilization Bands (RUB)	Non-user vs. Very High	-16.56	0.00	0.9210	<.0001
Resource Utilization Bands (RUB)	Healthy User vs. Very High	-2.03	0.13	<.0001	
Resource Utilization Bands (RUB)	Low Morbidity vs. Very High	-0.71	0.49	<.0001	
Resource Utilization Bands (RUB)	Moderate Morbidity vs. Very High	0.22	1.25	0.0003	
Resource Utilization Bands (RUB)	High Morbidity vs. Very High	0.30	1.35	<.0001	
Substance Abuse	No vs. Yes	0.13	1.14	0.0094	0.0094
Age	45-64 vs. Under 45	0.69	1.98	<.0001	<.0001
Age	65+ vs. Under 45	0.05	1.06	0.6917	
Gender	Female vs. Male	-0.03	0.97	0.5104	0.5104
Education	Unknown vs. More than High School	0.27	1.31	0.2490	0.3860
Education	Less than High School vs. More than High School	0.05	1.06	0.4511	
Education	High School Graduate vs. More than High School	-0.01	0.99	0.8849	
Language	Non-English vs. English	0.31	1.36	<.0001	<.0001
Metropolitan County	No vs. Yes	-0.19	0.82	<.0001	<.0001
Race / Ethnicity	Non-white vs. White	-0.21	0.81	<.0001	<.0001

Table 30. Risk adjustment model results for the Follow-up after hospitalization for mental illness within 7 days after discharge measure (FUH-AD-7).

Parameter	Level	Estimate	Odds Ratio	Category p-value	Variable p-value
Intercept		0.25	1.29	0.0542	
Developmental Disability	No vs. Yes	-0.06	0.94	0.5045	0.5045
Disability	No vs. Yes	0.00	1.00	0.9370	0.9370
Frailty	No vs. Yes	-0.07	0.93	0.1825	0.1825
Resource Utilization Bands (RUB)	Non-user vs. Very High	0.70	2.01	0.5693	0.0077
Resource Utilization Bands (RUB)	Healthy User vs. Very High	-0.77	0.47	0.3637	
Resource Utilization Bands (RUB)	Low Morbidity vs. Very High	-1.00	0.37	0.0015	
Resource Utilization Bands (RUB)	Moderate Morbidity vs. Very High	-0.15	0.86	0.0331	
Resource Utilization Bands (RUB)	High Morbidity vs. Very High	-0.02	0.98	0.7249	
Substance Abuse	No vs. Yes	-0.07	0.93	0.3523	0.3523
Age	45-64 vs. Under 45	-0.08	0.92	0.1951	0.1523
Age	65+ vs. Under 45	-1.19	0.31	0.1423	
Gender	Female vs. Male	0.15	1.17	0.0022	0.0022
Education	Unknown vs. More than High School	-0.19	0.83	0.4775	0.0245
Education	Less than High School vs. More than High School	-0.25	0.78	0.0032	
Education	High School Graduate vs. More than High School	-0.22	0.80	0.0054	
Language	Non-English vs. English	-0.10	0.91	0.4461	0.4461
Metropolitan County	No vs. Yes	-0.38	0.69	<.0001	<.0001
Race / Ethnicity	Non-white vs. White	0.05	1.06	0.3210	0.3210

Table 31. Risk adjustment model results for the Follow-up after hospitalization for mental illness within 30 days after discharge measure (FUH-AD-30).

Parameter	Level	Estimate	Odds Ratio	Category p-value	Variable p-value
Intercept		1.01	2.74	<.0001	
Developmental Disability	No vs. Yes	-0.19	0.83	0.0497	0.0497
Disability	No vs. Yes	-0.05	0.95	0.3920	0.3920
Frailty	No vs. Yes	-0.07	0.94	0.2446	0.2446
Resource Utilization Bands (RUB)	Non-user vs. Very High	0.19	1.20	0.8801	0.0002
Resource Utilization Bands (RUB	Healthy User vs. Very High	-1.22	0.30	0.1493	
Resource Utilization Bands (RUB	Low Morbidity vs. Very High	-0.97	0.38	0.0005	
Resource Utilization Bands (RUB	Moderate Morbidity vs. Very High	-0.21	0.81	0.0033	
Resource Utilization Bands (RUB	High Morbidity vs. Very High	-0.01	0.99	0.8207	
Substance Abuse	No vs. Yes	-0.11	0.90	0.1665	0.1665
Age	45-64 vs. Under 45	-0.03	0.97	0.6320	0.5692
Age	65+ vs. Under 45	-0.65	0.52	0.3383	
Gender	Female vs. Male	0.22	1.24	<.0001	<.0001
Education	Unknown vs. More than High School	-0.34	0.71	0.1984	0.0002
Education	Less than High School vs. More than High School	-0.40	0.67	<.0001	
Education	High School Graduate vs. More than High School	-0.28	0.75	0.0006	
Language	Non-English vs. English	-0.05	0.95	0.6942	0.6942
Metropolitan County	No vs. Yes	-0.13	0.88	0.0414	0.0414
Race / Ethnicity	Non-white vs. White	0.01	1.01	0.8187	0.8187

Table 32. Risk adjustment model results for the Adherence to antipsychotics for individuals with schizophrenia measure (SAA-AD).

Parameter	Level	Estimate	Odds Ratio	Category p-value	Variable p-value
Intercept		1.02	2.77	<.0001	
Developmental Disability	No vs. Yes	-0.45	0.64	<.0001	<.0001
Disability	No vs. Yes	-1.00	0.37	<.0001	<.0001
Frailty	No vs. Yes	0.38	1.47	<.0001	<.0001
Resource Utilization Bands (RUB)	Healthy User vs. Very High	-0.04	0.96	0.9600	0.8000
Resource Utilization Bands (RUB)	Low Morbidity vs. Very High	0.06	1.06	0.9406	
Resource Utilization Bands (RUB)	Moderate Morbidity vs. Very High	0.09	1.09	0.3368	
Resource Utilization Bands (RUB)	High Morbidity vs. Very High	0.00	1.00	0.9589	
Substance Abuse	No vs. Yes	0.50	1.66	<.0001	<.0001
Age	45-64 vs. Under 45	0.37	1.45	<.0001	<.0001
Gender	Female vs. Male	0.22	1.24	0.0012	0.0012
Education	Unknown vs. More than High School	0.10	1.10	0.8495	0.7159
Education	Less than High School vs. More than High School	-0.05	0.95	0.6886	
Education	High School Graduate vs. More than High School	-0.11	0.90	0.3591	
Language	Non-English vs. English	0.44	1.56	0.0008	0.0008
Metropolitan County	No vs. Yes	0.03	1.03	0.7536	0.7536
Race / Ethnicity	Non-white vs. White	-0.83	0.44	<.0001	<.0001

Table 33. Risk adjustment model results for the Antidepressant medication management acute phase treatment measure (AMM-AD_acute).

Parameter	Level	Estimate	Odds Ratio	Category p-value	Variable p-value
Intercept		0.69	1.99	<.0001	
Developmental Disability	No vs. Yes	-0.37	0.69	<.0001	<.0001
Disability	No vs. Yes	0.03	1.03	0.5196	0.5196
Frailty	No vs. Yes	-0.13	0.88	0.0031	0.0031
Resource Utilization Bands (RUB)	Non-user vs. Very High	-1.34	0.26	<.0001	<.0001
Resource Utilization Bands (RUB)	Healthy User vs. Very High	-0.35	0.70	0.0003	
Resource Utilization Bands (RUB)	Low Morbidity vs. Very High	-0.40	0.67	<.0001	
Resource Utilization Bands (RUB)	Moderate Morbidity vs. Very High	-0.20	0.82	<.0001	
Resource Utilization Bands (RUB)	High Morbidity vs. Very High	-0.21	0.81	<.0001	
Substance Abuse	No vs. Yes	-0.13	0.88	0.0001	0.0001
Age	45-64 vs. Under 45	0.38	1.47	<.0001	<.0001
	65+ vs. Under 45	0.73	2.08	0.0002	
Gender	Female vs. Male	-0.10	0.90	0.0013	0.0013
Education	Unknown vs. More than High School	0.10	1.10	0.5995	<.0001
Education	Less than High School vs. More than High School	-0.26	0.77	<.0001	
Education	High School Graduate vs. More than High School	-0.21	0.81	<.0001	
Language	Non-English vs. English	0.01	1.01	0.8515	0.8515
Metropolitan County	No vs. Yes	0.03	1.03	0.3632	0.3632
Race / Ethnicity	Non-white vs. White	-0.69	0.50	<.0001	<.0001

Table 34. Risk adjustment model results for the Antidepressant medication management continuation phase treatment measure (AMM-AD_cont).

Parameter	Level	Estimate	Odds Ratio	Category p-value	Variable p-value
Intercept		0.12	1.12	0.2951	
Developmental Disability	No vs. Yes	-0.45	0.64	<.0001	<.0001
Disability	No vs. Yes	-0.04	0.97	0.4571	0.4571
Frailty	No vs. Yes	-0.15	0.87	0.0025	0.0025
Resource Utilization Bands (RUB)	Non-user vs. Very High	-2.66	0.07	<.0001	<.0001
Resource Utilization Bands (RUB)	Healthy User vs. Very High	-0.67	0.51	<.0001	
Resource Utilization Bands (RUB)	Low Morbidity vs. Very High	-0.64	0.53	<.0001	
Resource Utilization Bands (RUB)	Moderate Morbidity vs. Very High	-0.32	0.72	<.0001	
Resource Utilization Bands (RUB)	High Morbidity vs. Very High	-0.28	0.76	<.0001	
Substance Abuse	No vs. Yes	-0.09	0.91	0.0180	0.0180
Age	45-64 vs. Under 45	0.61	1.84	<.0001	<.0001
Age	65+ vs. Under 45	0.93	2.54	<.0001	
Gender	Female vs. Male	-0.08	0.92	0.0226	0.0226
Education	Unknown vs. More than High School	0.22	1.25	0.2558	<.0001
Education	Less than High School vs. More than High School	-0.31	0.73	<.0001	
Education	High School Graduate vs. More than High School	-0.27	0.76	<.0001	
Language	Non-English vs. English	0.01	1.01	0.8881	0.8881
Metropolitan County	No vs. Yes	0.01	1.01	0.7721	0.7721
Race / Ethnicity	Non-white vs. White	-0.83	0.44	<.0001	<.0001

Table 35. Risk adjustment model results for the Initiation and engagement of alcohol and other drug dependence treatment within 14 days of the diagnosis measure (IET-AD-14).

Parameter	Level	Estimate	Odds Ratio	Category p-value	Variable p-value
Intercept		0.62	1.85	<.0001	
Developmental Disability	No vs. Yes	-0.16	0.85	0.0372	0.0372
Disability	No vs. Yes	-0.06	0.94	0.0554	0.0554
Frailty	No vs. Yes	-0.33	0.72	<.0001	<.0001
Mental Health	No vs. Yes	-0.39	0.68	0.0356	0.0356
Resource Utilization Bands (RUB)	Non-user vs. Very High	0.51	1.66	0.7228	<.0001
Resource Utilization Bands (RUB)	Healthy User vs. Very High	-2.45	0.09	<.0001	
Resource Utilization Bands (RUB)	Low Morbidity vs. Very High	-2.45	0.09	<.0001	
Resource Utilization Bands (RUB)	Moderate Morbidity vs. Very High	-1.18	0.31	<.0001	
Resource Utilization Bands (RUB)	High Morbidity vs. Very High	-0.49	0.62	<.0001	
Age	45-64 vs. Under 45	-0.23	0.79	<.0001	<.0001
Age	65+ vs. Under 45	-0.15	0.86	0.5350	
Gender	Female vs. Male	-0.04	0.96	0.1655	0.1655
Education	Unknown vs. More than High School	0.07	1.07	0.7061	0.1659
Education	Less than High School vs. More than High School	0.00	1.00	0.9617	
Education	High School Graduate vs. More than High School	-0.06	0.95	0.1733	
Language	Non-English vs. English	0.02	1.03	0.7862	0.7862
Metropolitan County	No vs. Yes	-0.01	0.99	0.6602	0.6602
Race / Ethnicity	Non-white vs. White	-0.12	0.89	<.0001	<.0001

Table 36. Risk adjustment model results for the Initiation and engagement of alcohol and other drug dependence treatment within 30 days of the initiation measure (IET-AD-30).

Parameter	Level	Estimate	Odds Ratio	Category p-value	Variable p-value
Intercept		-1.67	0.19	<.0001	
Developmental Disability	No vs. Yes	-0.23	0.79	0.1077	0.1077
Disability	No vs. Yes	0.08	1.08	0.2562	0.2562
Frailty	No vs. Yes	-0.38	0.69	<.0001	<.0001
Mental Health	No vs. Yes	-0.48	0.62	0.2920	0.2920
Resource Utilization Bands (RUB)	Non-user vs. Very High	-12.00	0.00	0.9923	<.0001
Resource Utilization Bands (RUB)	Healthy User vs. Very High	-12.03	0.00	0.9540	
Resource Utilization Bands (RUB)	Low Morbidity vs. Very High	-1.80	0.17	<.0001	
Resource Utilization Bands (RUB)	Moderate Morbidity vs. Very High	-0.67	0.51	<.0001	
Resource Utilization Bands (RUB)	High Morbidity vs. Very High	-0.30	0.74	<.0001	
Age	45-64 vs. Under 45	-0.29	0.75	<.0001	<.0001
Age	65+ vs. Under 45	-0.57	0.57	0.3357	
Gender	Female vs. Male	0.03	1.04	0.5216	0.5216
Education	Unknown vs. More than High School	0.01	1.01	0.9711	0.1455
Education	Less than High School vs. More than High School	-0.11	0.90	0.2148	
Education	High School Graduate vs. More than High School	-0.17	0.84	0.0270	
Language	Non-English vs. English	-0.52	0.60	0.0385	0.0385
Metropolitan County	No vs. Yes	-0.06	0.94	0.3213	0.3213
Race / Ethnicity	Non-white vs. White	-0.39	0.67	<.0001	<.0001

Table 37. Risk adjustment model results for the Diabetes short-term complications admission rate measure (PQI01-AD).

Parameter	Level	Estimate	Odds Ratio	Category p-value	Variable p-value
Intercept		-3.89	0.02	<.0001	
Resource Utilization Bands (RUB)	Non-user vs. Very High	-18.53	0.00	0.9477	<.0001
Resource Utilization Bands (RUB)	Healthy User vs. Very High	-6.70	0.00	<.0001	
Resource Utilization Bands (RUB)	Low Morbidity vs. Very High	-7.08	0.00	<.0001	
Resource Utilization Bands (RUB)	Moderate Morbidity vs. Very High	-4.48	0.01	<.0001	
Resource Utilization Bands (RUB)	High Morbidity vs. Very High	-2.12	0.12	<.0001	
Age	45-64 vs. Under 45	-0.83	0.44	<.0001	<.0001
Age	65+ vs. Under 45	-1.37	0.25	0.0198	
Gender	Female vs. Male	-0.66	0.52	<.0001	<.0001
Education	Unknown vs. More than High School	-0.13	0.88	0.8021	0.2027
Education	Less than High School vs. More than High School	0.26	1.29	0.1140	
Education	High School Graduate vs. More than High School	0.30	1.36	0.0454	
Language	Non-English vs. English	-0.75	0.47	0.0005	0.0005
Metropolitan County	No vs. Yes	-0.05	0.95	0.6295	0.6295
Race / Ethnicity	Non-white vs. White	0.32	1.38	0.0004	0.0004

Table 38. Risk adjustment model results for the Chronic obstructive pulmonary disease (COPD) or asthma in older adults admission rate measure (PQI05-AD).

Parameter	Level	Estimate	Odds Ratio	Category p-value	Variable p-value
Intercept		-4.77	0.01	<.0001	
Resource Utilization Bands (RUB)	Non-user vs. Very High	-17.10	0.00	0.9536	<.0001
Resource Utilization Bands (RUB)	Healthy User vs. Very High	-17.12	0.00	0.9488	
Resource Utilization Bands (RUB)	Low Morbidity vs. Very High	-5.04	0.01	<.0001	
Resource Utilization Bands (RUB)	Moderate Morbidity vs. Very High	-3.54	0.03	<.0001	
Resource Utilization Bands (RUB)	High Morbidity vs. Very High	-1.54	0.22	<.0001	
Disability	No vs. Yes	-0.61	0.55	<.0001	<.0001
Age	45-64 vs. Under 45	0.76	2.13	<.0001	<.0001
Age	65+ vs. Under 45	1.65	5.23	<.0001	
Gender	Female vs. Male	0.41	1.51	<.0001	<.0001
Education	Unknown vs. More than High School	0.37	1.44	0.3381	0.0003
Education	Less than High School vs. More than High School	0.64	1.90	<.0001	
Education	High School Graduate vs. More than High School	0.50	1.65	0.0003	
Language	Non-English vs. English	-0.61	0.54	0.0001	0.0001
Metropolitan County	No vs. Yes	-0.11	0.90	0.2662	0.2662
Race / Ethnicity	Non-white vs. White	0.15	1.16	0.0965	0.0965

Table 39. Risk adjustment model results for the Postpartum care rate measure (PPC-AD).

Parameter	Level	Estimate	Odds Ratio	Category p-value	Variable p-value
Intercept		-0.35	0.71	0.2794	
Resource Utilization Bands (RUB)	Non-user vs. Very High	-1.99	0.14	<.0001	<.0001
Resource Utilization Bands (RUB)	Healthy User vs. Very High	-0.07	0.93	0.8303	
Resource Utilization Bands (RUB)	Low Morbidity vs. Very High	-0.10	0.91	0.7626	
Resource Utilization Bands (RUB)	Moderate Morbidity vs. Very High	-0.12	0.89	0.7143	
Resource Utilization Bands (RUB)	High Morbidity vs. Very High	-0.21	0.81	0.5074	
Age	45-64 vs. Under 45	-0.30	0.74	0.2163	0.2163
Education	Unknown vs. More than High School	0.25	1.29	0.0195	0.0089
Education	Less than High School vs. More than High School	0.01	1.01	0.8381	
Education	High School Graduate vs. More than High School	-0.05	0.95	0.2886	
Language	Non-English vs. English	0.69	1.99	<.0001	<.0001
Metropolitan County	No vs. Yes	0.31	1.36	<.0001	<.0001
Race / Ethnicity	Non-white vs. White	-0.13	0.88	<.0001	<.0001

C. Risk Adjusted Rates

Table 40. Breast cancer screening (BCS-AD).

MCO	Denominator (Members)	Observed (Members)	Expected (Members)	Observed Rate	Expected Rate	Risk Adjusted Rate	Lower 95% CI	Upper 95% CI
FFS	4,710	2,584	2,662	54.9%	56.5%	58.0%	56.5%	59.5%
MCO 1	368	227	220	61.7%	59.7%	61.8%	57.0%	66.6%
MCO 2	3,483	2,242	2,197	64.4%	63.1%	61.0%	59.5%	62.4%
MCO 3	158	108	103	68.4%	64.9%	63.0%	56.3%	69.6%
MCO 4	447	241	278	53.9%	62.1%	51.9%	47.7%	56.0%
MCO 5	4,175	2,460	2,512	58.9%	60.2%	58.5%	57.1%	59.9%
MCO 6	5,508	3,303	3,268	60.0%	59.3%	60.4%	59.1%	61.7%
MCO 7	1,712	1,128	1,050	65.9%	61.3%	64.2%	62.1%	66.4%
MCO 8	153	94	100	61.4%	65.3%	56.3%	49.6%	63.0%
MCO 9	259	143	142	55.2%	54.9%	60.0%	53.6%	66.5%
Total	20,973	12,530		59.7%				

Figure 4. Breast cancer screening (BCS-AD).

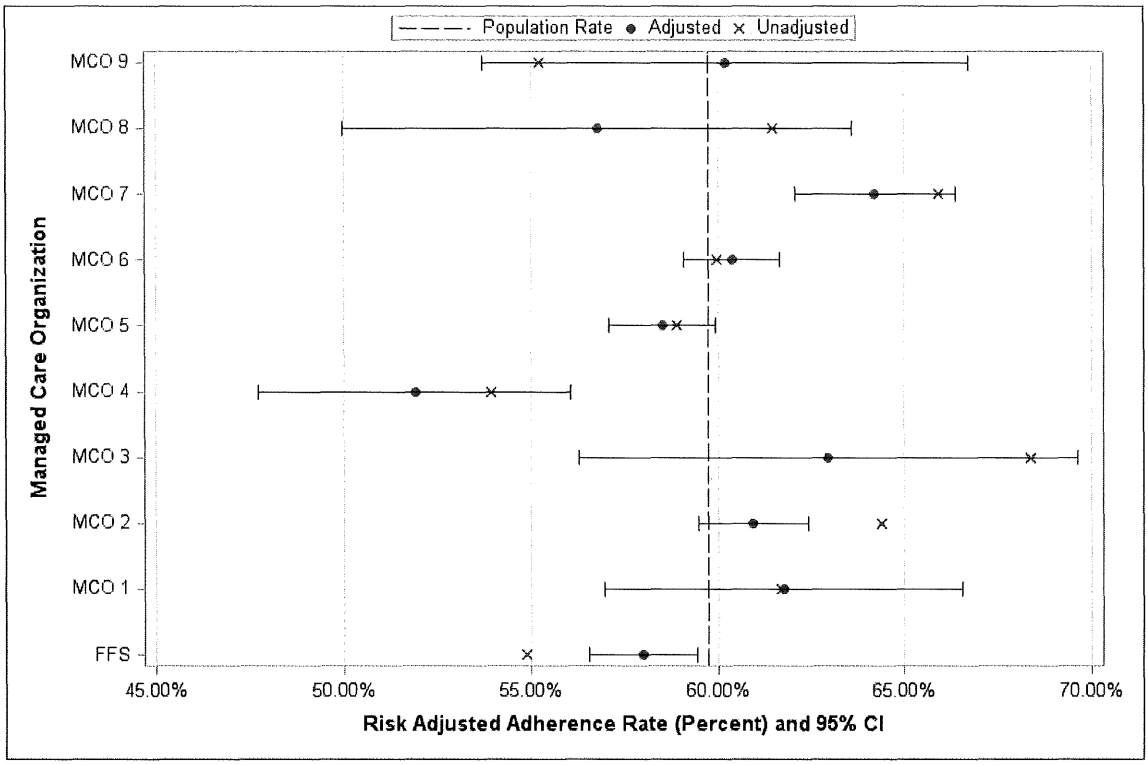


Table 41. Cervical cancer screening (CCS-AD).

MCO	Denominator (Members)	Observed (Members)	Expected (Members)	Observed Rate	Expected Rate	Risk Adjusted Rate	Lower 95% CI	Upper 95% CI
FFS	16,106	8,994	9,657	55.8%	60.0%	61.5%	60.7%	62.3%
MCO 1	2,157	1,476	1,445	68.4%	67.0%	67.4%	65.6%	69.3%
MCO 2	13,885	9,579	9,371	69.0%	67.5%	67.5%	66.7%	68.2%
MCO 3	664	453	441	68.2%	66.5%	67.7%	64.3%	71.2%
MCO 4	2,241	1,459	1,471	65.1%	65.6%	65.5%	63.6%	67.4%
MCO 5	14,856	9,889	9,823	66.6%	66.1%	66.4%	65.7%	67.2%
MCO 6	21,929	14,967	14,826	68.3%	67.6%	66.6%	66.1%	67.2%
MCO 7	8,444	6,257	6,021	74.1%	71.3%	68.6%	67.7%	69.5%
MCO 8	356	198	240	55.6%	67.5%	54.4%	49.8%	59.0%
MCO 9	541	307	281	56.8%	52.0%	72.1%	66.9%	77.2%
Total	81,179	53,579		66.0%				

Figure 5. Cervical cancer screening (CCS-AD).

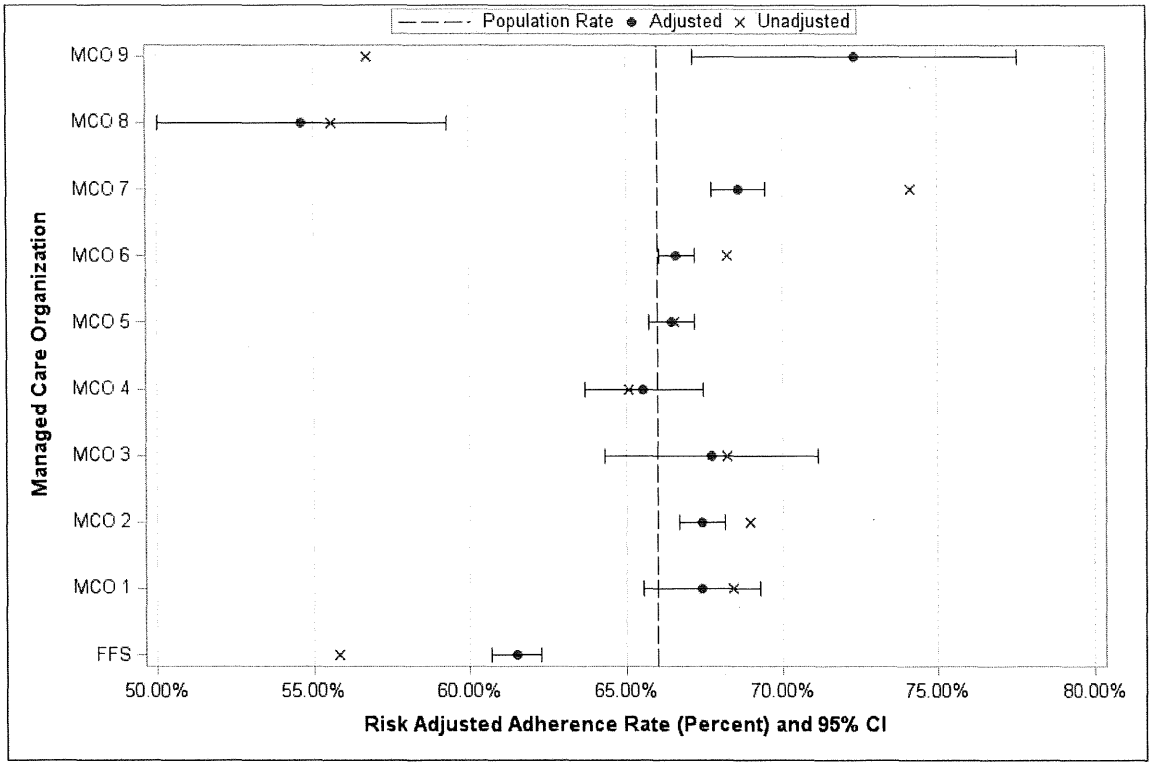


Table 42. Chlamydia screening in women (CHL-AD).

MCO	Denominator (Members)	Observed (Members)	Expected (Members)	Observed Rate	Expected Rate	Risk Adjusted Rate	Lower 95% CI	Upper 95% CI
FFS	5,937	2,824	2,982	47.6%	50.2%	54.9%	53.5%	56.3%
MCO 1	457	267	240	58.4%	52.6%	64.4%	59.5%	69.3%
MCO 2	2,363	1,334	1,356	56.5%	57.4%	57.1%	55.1%	59.0%
MCO 3	133	69	71	51.9%	53.4%	56.4%	47.3%	65.5%
MCO 4	487	211	258	43.3%	52.9%	47.5%	42.8%	52.3%
MCO 5	2,368	1,572	1,520	66.4%	64.2%	60.0%	58.3%	61.7%
MCO 6	4,013	2,626	2,572	65.4%	64.1%	59.2%	57.9%	60.5%
MCO 7	1,540	1,112	1,029	72.2%	66.8%	62.7%	60.7%	64.7%
MCO 8	49	40	35	81.6%	72.1%	65.7%	55.8%	75.7%
MCO 9	38	29	21	76.3%	54.8%	80.7%	64.5%	97.0%
Total	17,385	10,084		58.0%				

Figure 6. Chlamydia screening in women (CHL-AD).

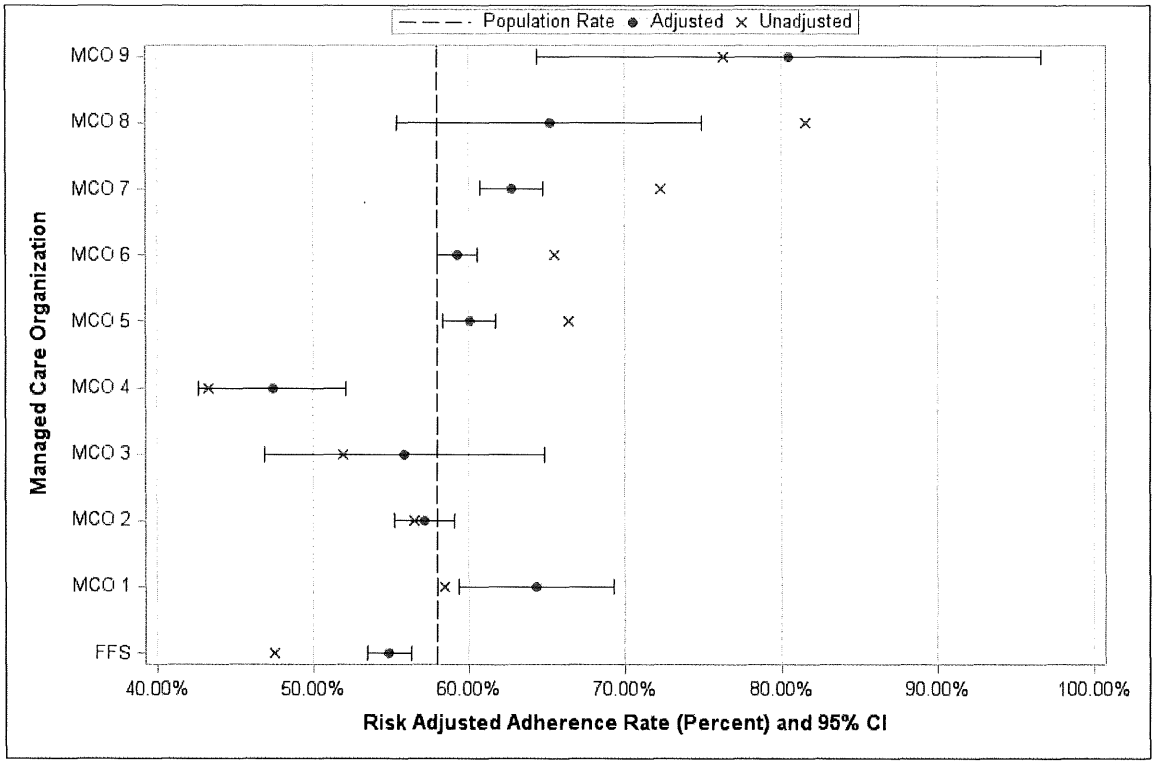


Table 43. Annual monitoring for enrollees on angiotensin converting enzyme (ACE) inhibitors or angiotensin receptor blockers (ARB) (MPM-AD-R1).

MCO	Denominator (Members)	Observed (Members)	Expected (Members)	Observed Rate	Expected Rate	Risk Adjusted Rate	Lower 95% CI	Upper 95% CI
FFS	4,460	3,863	3,911	86.6%	87.7%	85.6%	84.7%	86.5%
MCO 1	484	408	404	84.3%	83.5%	87.5%	84.2%	90.7%
MCO 2	3,679	3,115	3,108	84.7%	84.5%	86.8%	85.7%	88.0%
MCO 3	181	156	152	86.2%	84.2%	88.7%	83.5%	94.0%
MCO 4	571	459	484	80.4%	84.7%	82.2%	79.3%	85.1%
MCO 5	3,991	3,494	3,482	87.6%	87.3%	86.9%	85.9%	87.9%
MCO 6	5,532	4,869	4,834	88.0%	87.4%	87.2%	86.4%	88.1%
MCO 7	2,058	1,781	1,762	86.5%	85.6%	87.6%	86.1%	89.0%
MCO 8	315	262	276	83.2%	87.6%	82.3%	78.8%	85.7%
MCO 9	261	242	236	92.7%	90.6%	88.7%	85.4%	92.0%
Total	21,532	18,649		86.6%				

Figure 7. Annual monitoring for enrollees on angiotensin converting enzyme (ACE) inhibitors or angiotensin receptor blockers (ARB) (MPM-AD-R1).

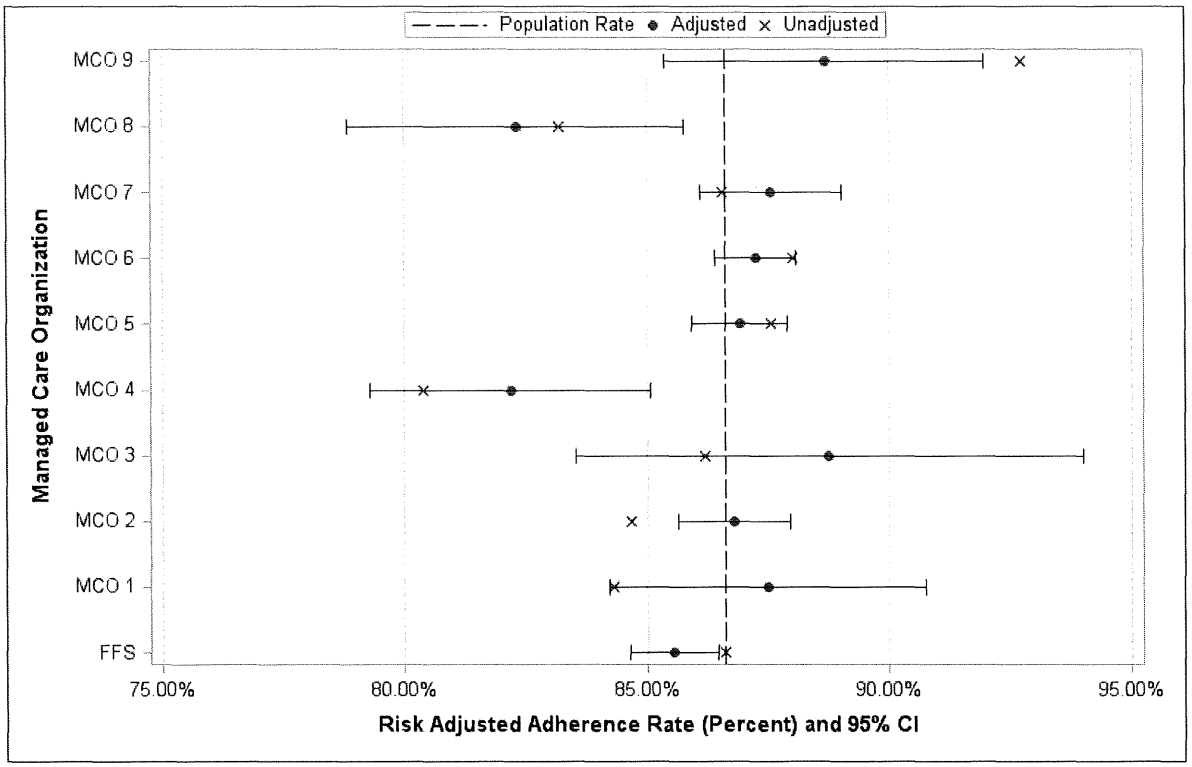


Table 44. Annual monitoring for enrollees on diuretics (MPM-AD-R3).

MCO	Denominator (Members)	Observed (Members)	Expected (Members)	Observed Rate	Expected Rate	Risk Adjusted Rate	Lower 95% CI	Upper 95% CI
FFS	3,533	3,117	3,136	88.2%	88.8%	86.4%	85.5%	87.4%
MCO 1	367	310	311	84.5%	84.8%	86.6%	83.0%	90.1%
MCO 2	2,707	2,295	2,293	84.8%	84.7%	87.0%	85.7%	88.4%
MCO 3	154	129	129	83.8%	83.9%	86.8%	81.0%	92.6%
MCO 4	400	338	340	84.5%	85.0%	86.5%	83.1%	89.9%
MCO 5	3,226	2,809	2,817	87.1%	87.3%	86.7%	85.6%	87.8%
MCO 6	4,043	3,549	3,536	87.8%	87.5%	87.3%	86.3%	88.2%
MCO 7	1,659	1,441	1,418	86.9%	85.5%	88.4%	86.7%	90.0%
MCO 8	276	237	242	85.9%	87.8%	85.1%	81.4%	88.7%
MCO 9	213	190	192	89.2%	90.2%	86.0%	82.3%	89.7%
Total	16,578	14,415		87.0%				

Figure 8. Annual monitoring for enrollees on diuretics (MPM-AD-R3).

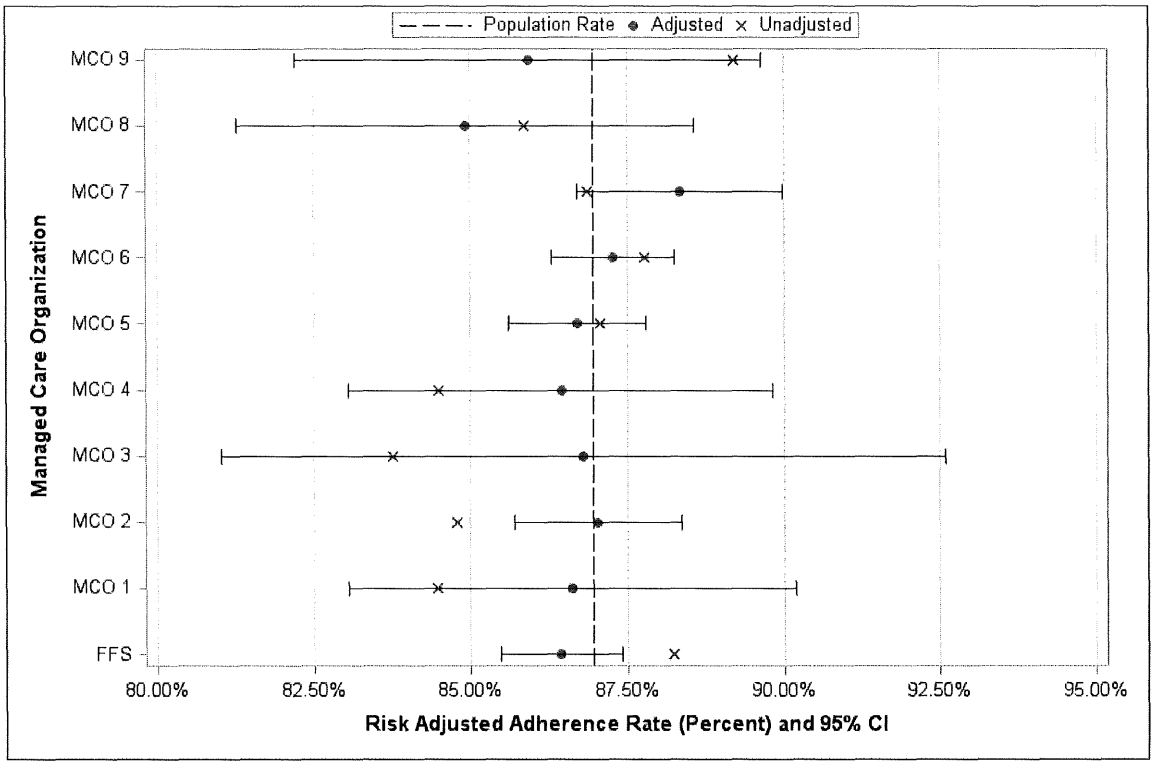


Table 45. Annual monitoring for enrollees on anticonvulsants (MPM-AD-R4).

MCO	Denominator (Members)	Observed (Members)	Expected (Members)	Observed Rate	Expected Rate	Risk Adjusted Rate	Lower 95% CI	Upper 95% CI
FFS	1,993	1,420	1,419	71.3%	71.2%	70.1%	68.2%	72.0%
MCO 1	88	55	58	62.5%	65.6%	66.8%	56.5%	77.2%
MCO 2	313	211	201	67.4%	64.3%	73.5%	67.9%	79.1%
MCO 3	23	17	15	73.9%	64.5%	80.3%	59.6%	100.0%
MCO 4	83	53	54	63.9%	64.8%	69.1%	58.3%	79.9%
MCO 5	712	500	509	70.2%	71.4%	68.9%	65.8%	72.1%
MCO 6	849	583	592	68.7%	69.7%	69.1%	66.0%	72.1%
MCO 7	181	135	122	74.6%	67.3%	77.7%	70.8%	84.5%
MCO 8	32	22	24	68.8%	74.5%	64.7%	50.9%	78.5%
MCO 9	52	36	39	69.2%	75.2%	64.5%	53.9%	75.2%
Total	4,326	3,032		70.1%				

Figure 9. Annual monitoring for enrollees on anticonvulsants (MPM-AD-R4).

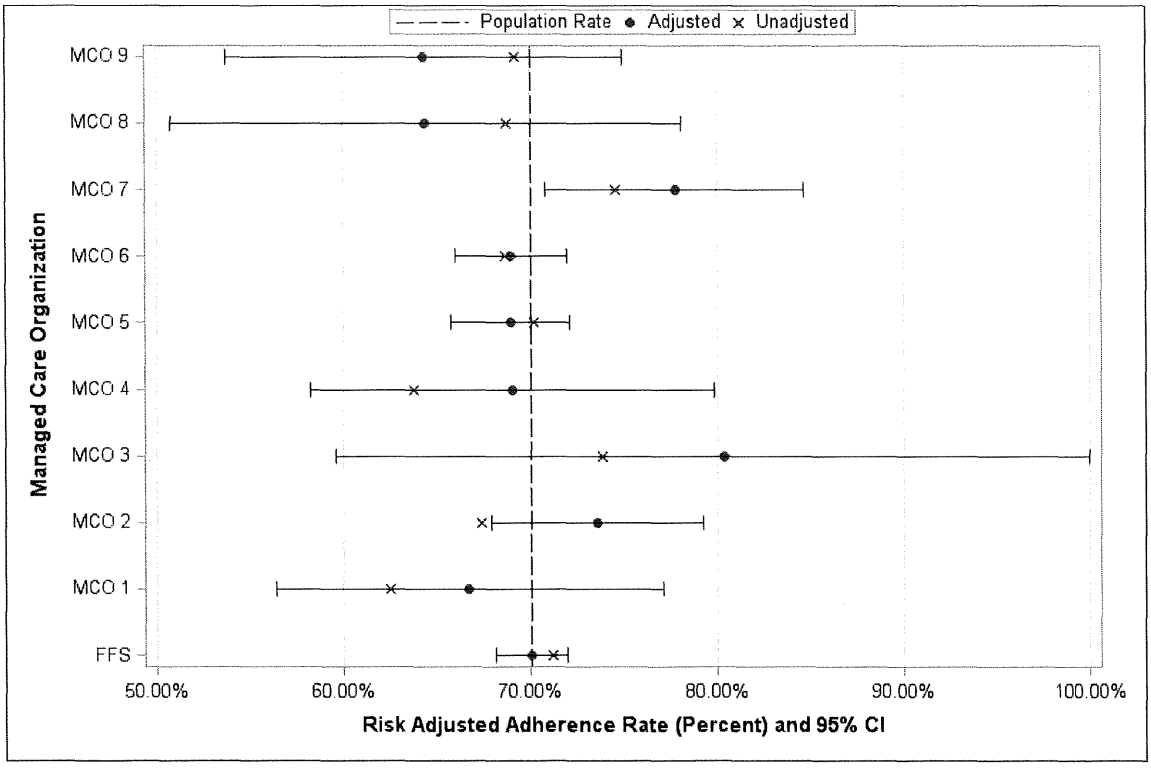


Table 46. Comprehensive diabetes care: Hemoglobin A1c testing (HA1C-AD).

MCO	Denominator (Members)	Observed (Members)	Expected (Members)	Observed Rate	Expected Rate	Risk Adjusted Rate	Lower 95% CI	Upper 95% CI
FFS	5,794	4,803	5,006	82.9%	86.4%	84.5%	83.6%	85.3%
MCO 1	467	424	406	90.8%	86.9%	92.0%	89.0%	94.9%
MCO 2	3,168	2,859	2,819	90.3%	89.0%	89.3%	88.2%	90.3%
MCO 3	173	146	152	84.4%	87.7%	84.7%	80.0%	89.4%
MCO 4	594	488	514	82.2%	86.5%	83.7%	81.0%	86.3%
MCO 5	4,363	3,889	3,854	89.1%	88.3%	88.8%	87.9%	89.8%
MCO 6	6,366	5,733	5,647	90.1%	88.7%	89.4%	88.6%	90.1%
MCO 7	2,238	2,054	1,995	91.8%	89.1%	90.6%	89.4%	91.9%
MCO 8	317	280	282	88.3%	89.0%	87.4%	84.1%	90.7%
MCO 9	287	248	249	86.4%	86.7%	87.7%	83.8%	91.6%
Total	23,767	20,924		88.0%				

Figure 10. Comprehensive diabetes care: Hemoglobin A1c testing (HA1C-AD).

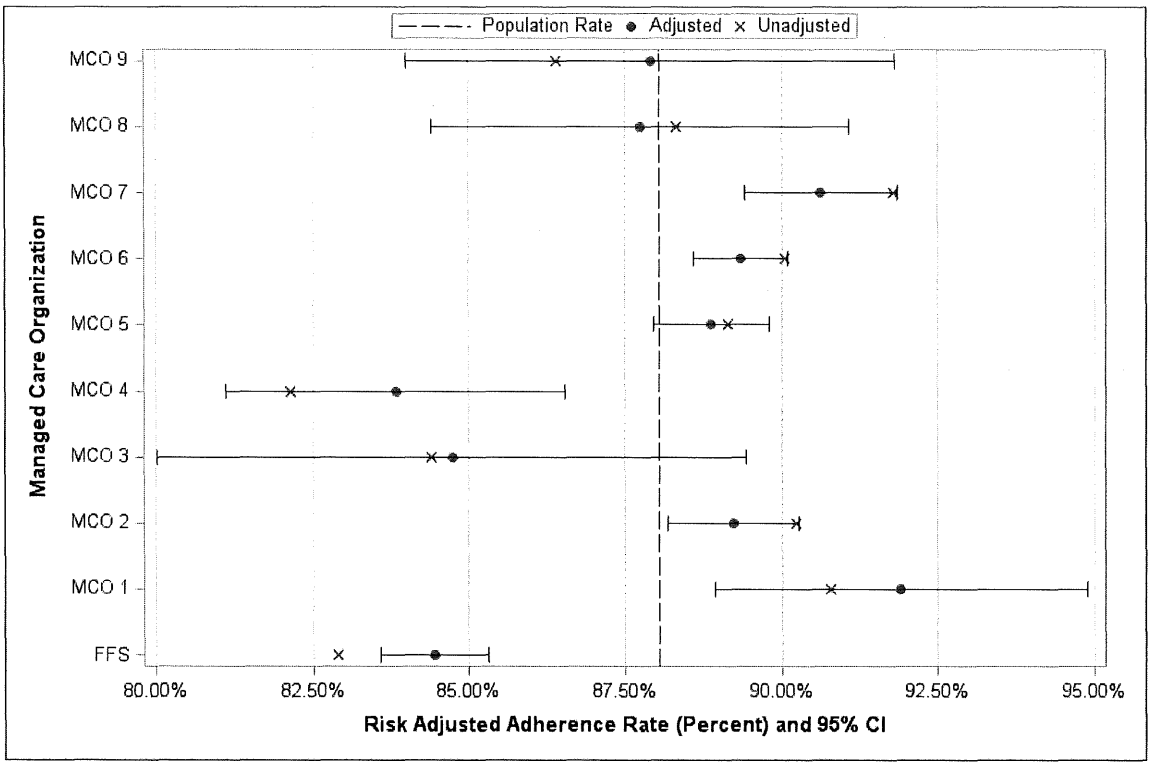


Table 47. Follow-up after hospitalization for mental illness within 7 days after discharge (FUH-AD-7).

MCO	Denominator (Members)	Observed (Members)	Expected (Members)	Observed Rate	Expected Rate	Risk Adjusted Rate	Lower 95% CI	Upper 95% CI
FFS	2,387	1,087	1,103	45.5%	46.2%	46.1%	44.1%	48.1%
MCO 1	105	25	43	23.8%	41.2%	27.0%	16.4%	37.6%
MCO 2	650	325	289	50.0%	44.5%	52.5%	48.5%	56.5%
MCO 3	25	6	10	24.0%	39.9%	28.1%	5.7%	50.6%
MCO 4	109	47	44	43.1%	40.1%	50.3%	39.6%	60.9%
MCO 5	896	382	429	42.6%	47.9%	41.6%	38.5%	44.8%
MCO 6	1,607	768	767	47.8%	47.7%	46.8%	44.4%	49.2%
MCO 7	565	336	275	59.5%	48.7%	57.0%	53.1%	61.0%
MCO 8	173	73	83	42.2%	47.8%	41.2%	34.0%	48.5%
MCO 9	68	28	33	41.2%	48.3%	39.9%	28.4%	51.3%
Total	6,585	3,077		46.7%				

Figure 11. Follow-up after hospitalization for mental illness within 7 days after discharge (FUH-AD-7).

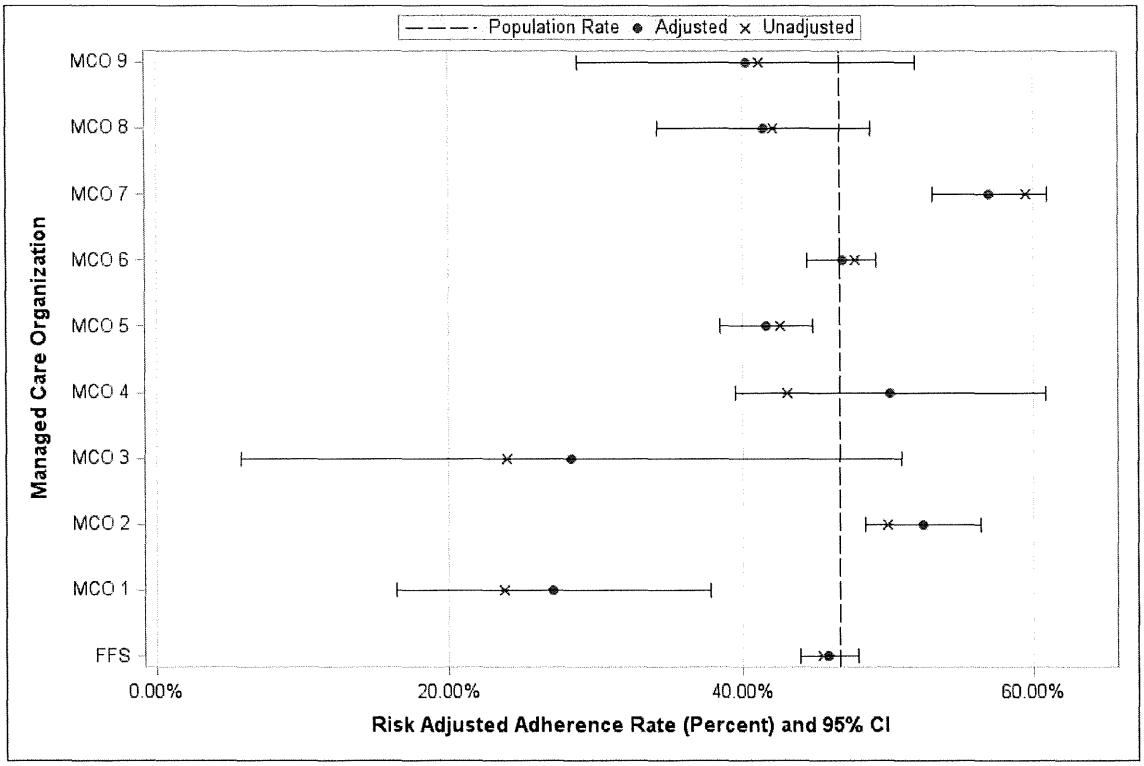


Table 48. Follow-up after hospitalization for mental illness within 30 days after discharge (FUH-AD-30).

MCO	Denominator (Members)	Observed (Members)	Expected (Members)	Observed Rate	Expected Rate	Risk Adjusted Rate	Lower 95% CI	Upper 95% CI
FFS	2,387	1,410	1,451	59.1%	60.8%	59.2%	57.3%	61.1%
MCO 1	105	59	62	56.2%	58.7%	58.3%	48.6%	68.0%
MCO 2	650	419	389	64.5%	59.9%	65.6%	61.8%	69.4%
MCO 3	25	14	15	56.0%	58.6%	58.2%	38.2%	78.2%
MCO 4	109	66	64	60.6%	58.5%	63.1%	53.5%	72.7%
MCO 5	896	509	553	56.8%	61.7%	56.1%	53.0%	59.2%
MCO 6	1,607	995	984	61.9%	61.2%	61.6%	59.2%	64.0%
MCO 7	565	398	346	70.4%	61.3%	70.0%	66.1%	74.0%
MCO 8	173	99	104	57.2%	60.0%	58.1%	50.7%	65.5%
MCO 9	68	43	42	63.2%	62.3%	61.8%	50.6%	73.0%
Total	6,585	4,012		60.9%				

Figure 12. Follow-up after hospitalization for mental illness within 30 days after discharge (FUH-AD-30).

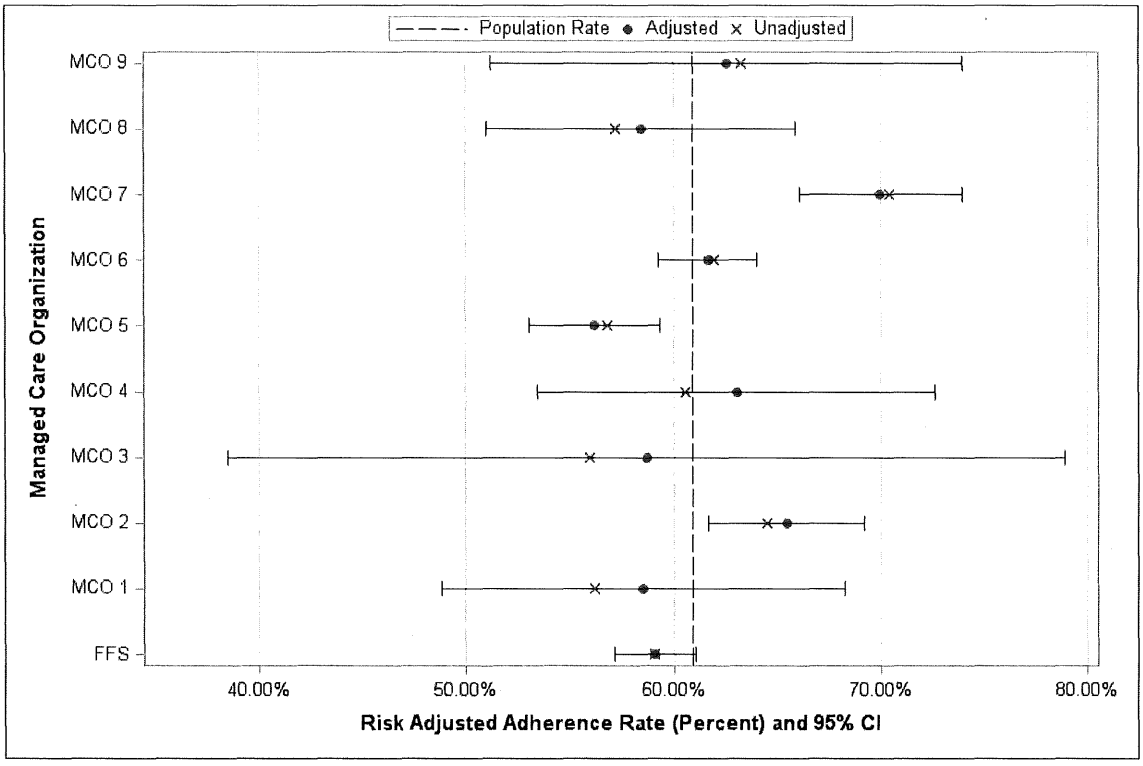


Table 49. Adherence to antipsychotics for individuals with schizophrenia (SAA-AD).

MCO	Denominator (Members)	Observed (Members)	Expected (Members)	Observed Rate	Expected Rate	Risk Adjusted Rate	Lower 95% CI	Upper 95% CI
FFS	2,271	1,645	1,601	72.4%	70.5%	67.2%	65.5%	68.8%
MCO 1	60	46	42	76.7%	69.7%	71.9%	61.7%	82.2%
MCO 2	144	77	77	53.5%	53.7%	65.1%	55.6%	74.6%
MCO 3	*	*	*	*	*	*	*	*
MCO 4	55	37	35	67.3%	64.3%	68.4%	56.3%	80.5%
MCO 5	816	554	534	67.9%	65.5%	67.8%	64.7%	70.8%
MCO 6	1,100	627	673	57.0%	61.2%	60.9%	58.0%	63.8%
MCO 7	193	94	90	48.7%	46.7%	68.1%	58.7%	77.6%
MCO 8	80	24	33	30.0%	41.5%	47.2%	30.8%	63.6%
MCO 9	181	100	118	55.3%	65.3%	55.3%	48.6%	62.0%
Total	4,908	3,208		65.4%				

Note: * denotes values masked due to small numbers (n < 5).

Figure 13. Adherence to antipsychotics for individuals with schizophrenia (SAA-AD).

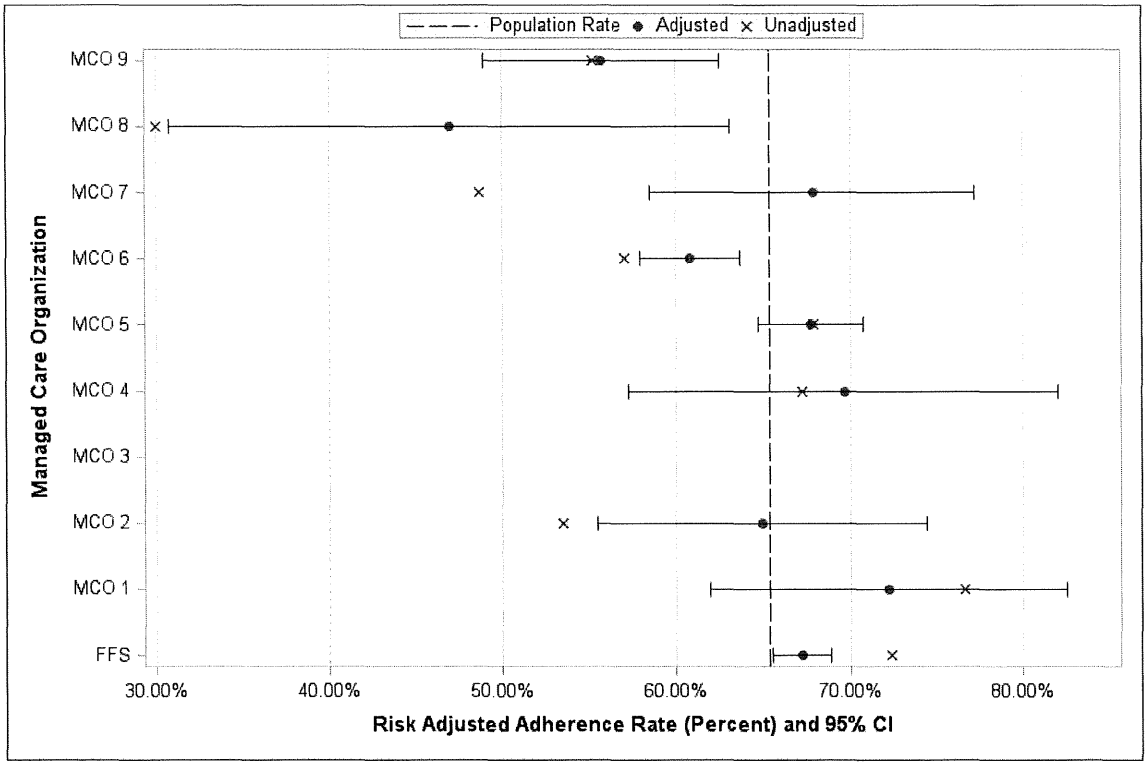


Table 50. Antidepressant medication management acute phase treatment (AMM-AD_acute).

MCO	Denominator (Members)	Observed (Members)	Expected (Members)	Observed Rate	Expected Rate	Risk Adjusted Rate	Lower 95% CI	Upper 95% CI
FFS	4,254	1,747	1,725	41.1%	40.6%	40.9%	39.4%	42.3%
MCO 1	524	212	224	40.5%	42.8%	38.1%	34.2%	42.0%
MCO 2	3,003	1,319	1,282	43.9%	42.7%	41.5%	39.9%	43.2%
MCO 3	158	57	68	36.1%	43.2%	33.7%	26.6%	40.8%
MCO 4	592	251	253	42.4%	42.8%	40.0%	36.3%	43.7%
MCO 5	3,518	1,346	1,401	38.3%	39.8%	38.8%	37.2%	40.4%
MCO 6	5,357	2,119	2,104	39.6%	39.3%	40.6%	39.3%	42.0%
MCO 7	2,137	847	838	39.6%	39.2%	40.8%	38.7%	42.8%
MCO 8	382	148	146	38.7%	38.2%	41.0%	35.9%	46.0%
MCO 9	140	48	52	34.3%	37.1%	37.3%	28.7%	45.8%
Total	20,065	8,094		40.3%				

Figure 14. Antidepressant medication management acute phase treatment (AMM-AD_acute).

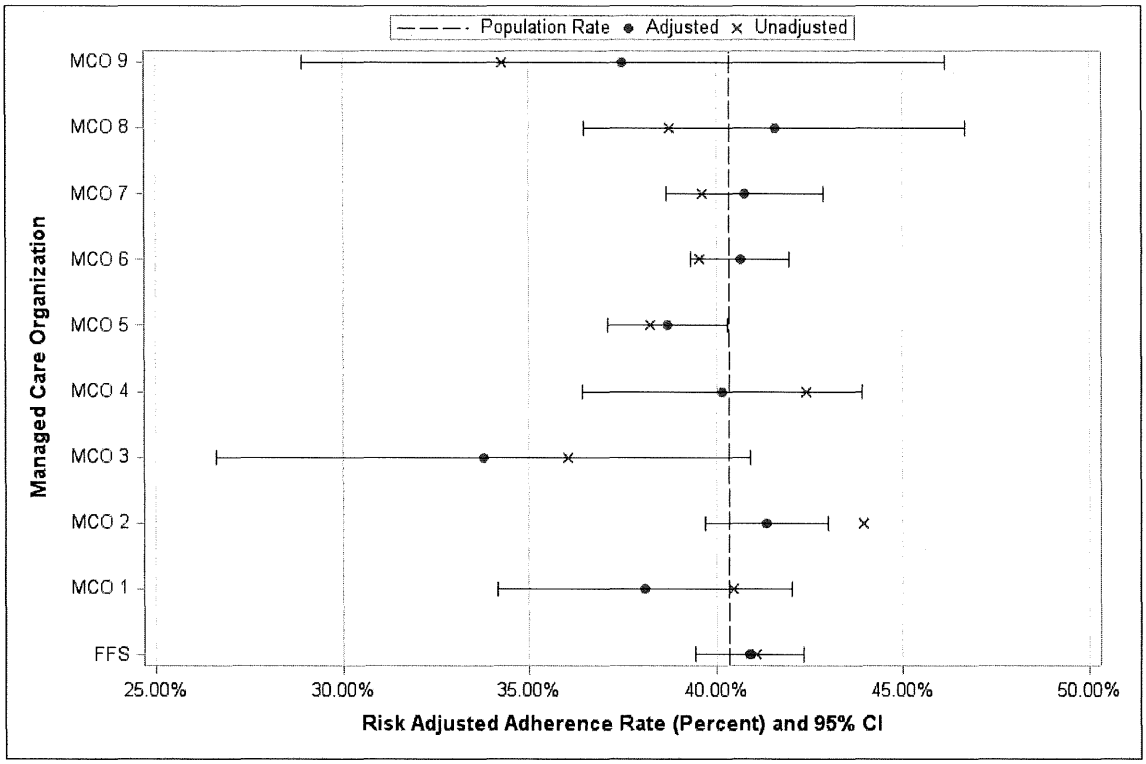


Table 51. Antidepressant medication management continuation phase treatment (AMM-AD_cont).

MCO	Denominator (Members)	Observed (Members)	Expected (Members)	Observed Rate	Expected Rate	Risk Adjusted Rate	Lower 95% CI	Upper 95% CI
FFS	4,254	1,000	1,043	23.5%	24.5%	22.9%	21.7%	24.1%
MCO 1	524	121	132	23.1%	25.1%	21.9%	18.5%	25.4%
MCO 2	3,003	752	761	25.0%	25.3%	23.6%	22.2%	25.0%
MCO 3	158	35	40	22.2%	25.2%	21.0%	14.7%	27.3%
MCO 4	592	139	150	23.5%	25.3%	22.2%	18.9%	25.4%
MCO 5	3,518	852	832	24.2%	23.6%	24.4%	23.1%	25.8%
MCO 6	5,357	1,276	1,230	23.8%	23.0%	24.8%	23.6%	25.9%
MCO 7	2,137	514	485	24.1%	22.7%	25.3%	23.5%	27.1%
MCO 8	382	70	85	18.3%	22.2%	19.7%	15.4%	24.1%
MCO 9	140	28	32	20.0%	22.6%	21.1%	14.0%	28.2%
Total	20,065	4,787		23.9%				

Figure 15. Antidepressant medication management continuation phase treatment (AMM-AD_cont).

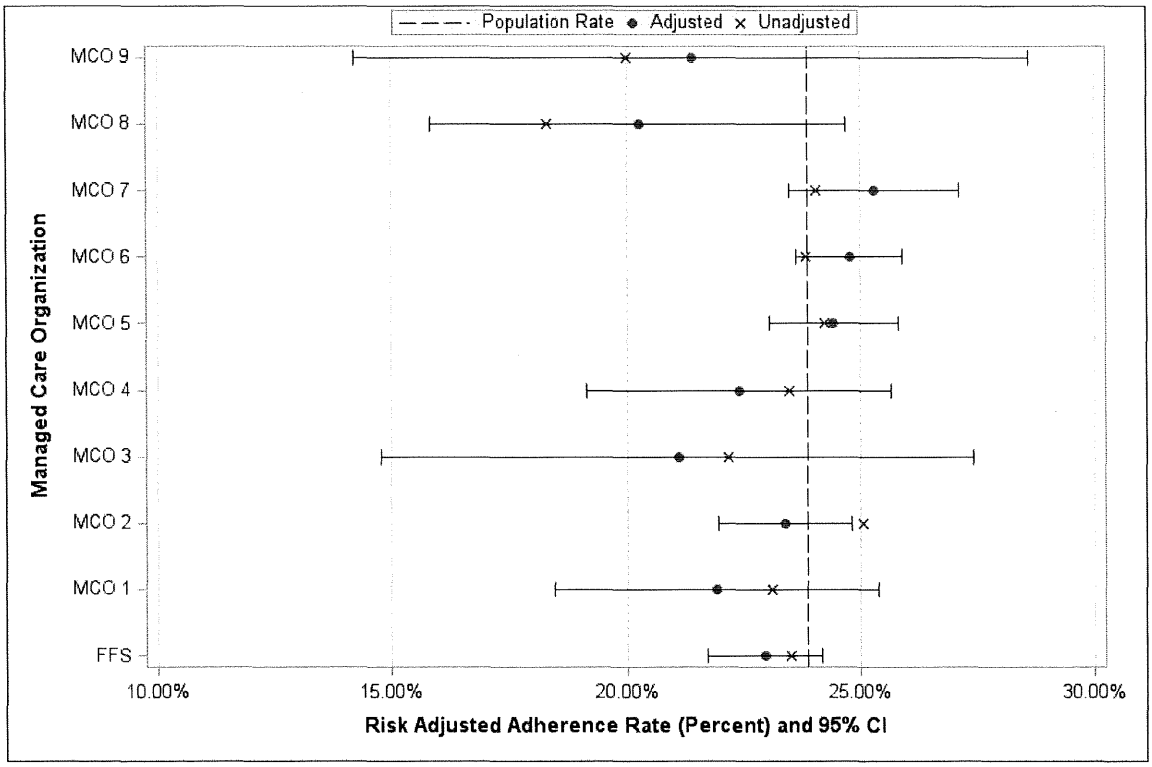


Table 52. Initiation and engagement of alcohol and other drug dependence treatment within 14 days of the diagnosis (IET-AD-14).

MCO	Denominator (Members)	Observed (Members)	Expected (Members)	Observed Rate	Expected Rate	Risk Adjusted Rate	Lower 95% CI	Upper 95% CI
FFS	7,057	2,634	2,532	37.3%	35.9%	35.8%	34.7%	36.8%
MCO 1	610	179	196	29.3%	32.1%	31.4%	27.6%	35.2%
MCO 2	3,332	1,051	1,092	31.5%	32.8%	33.1%	31.5%	34.7%
MCO 3	239	74	80	31.0%	33.3%	31.9%	25.9%	37.9%
MCO 4	762	217	255	28.5%	33.4%	29.3%	26.0%	32.6%
MCO 5	5,139	1,748	1,784	34.0%	34.7%	33.7%	32.4%	34.9%
MCO 6	7,255	2,619	2,510	36.1%	34.6%	35.9%	34.8%	36.9%
MCO 7	2,962	912	956	30.8%	32.3%	32.8%	31.0%	34.5%
MCO 8	1,003	301	327	30.0%	32.6%	31.7%	28.7%	34.6%
MCO 9	352	131	135	37.2%	38.4%	33.3%	28.9%	37.7%
Total	28,711	9,866		34.4%				

Figure 16. Initiation and engagement of alcohol and other drug dependence treatment within 14 days of the diagnosis (IET-AD-14).

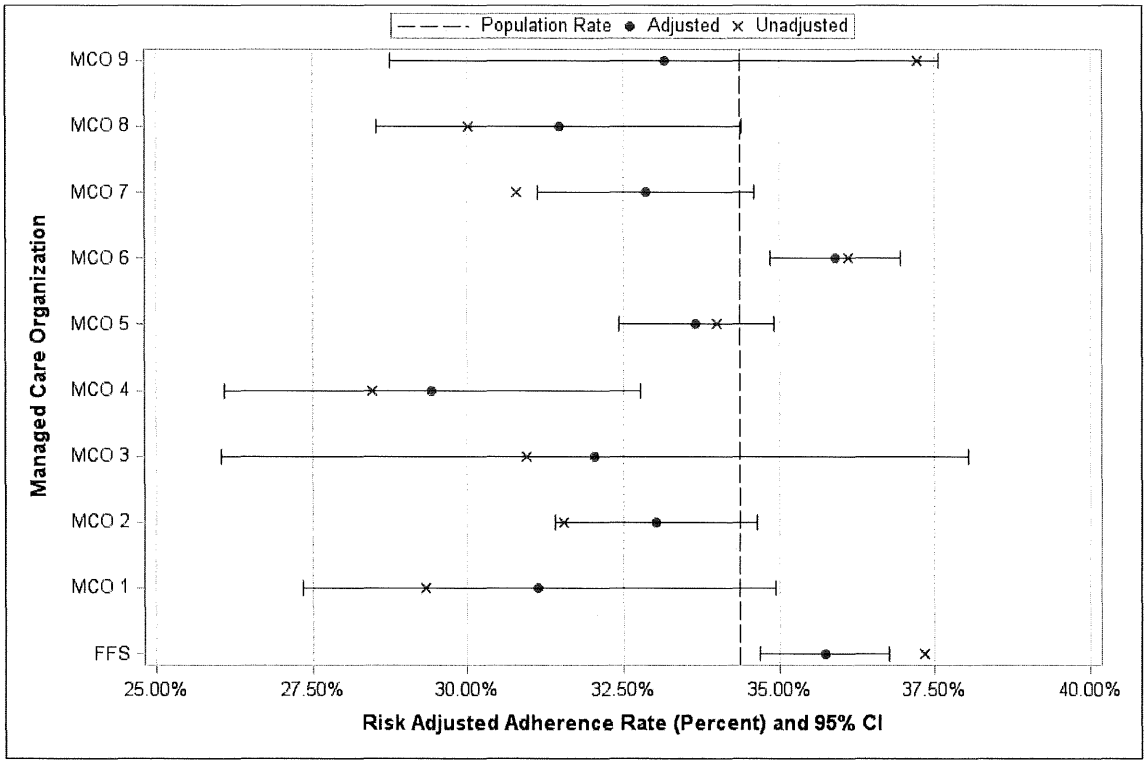


Table 53. Initiation and engagement of alcohol and other drug dependence treatment within 30 days of the initiation visit (IET-AD-30).

MCO	Denominator (Members)	Observed (Members)	Expected (Members)	Observed Rate	Expected Rate	Risk Adjusted Rate	Lower 95% CI	Upper 95% CI
FFS	7,057	377	388	5.3%	5.5%	5.3%	4.8%	5.8%
MCO 1	610	34	33	5.6%	5.4%	5.7%	3.9%	7.5%
MCO 2	3,332	189	188	5.7%	5.6%	5.5%	4.8%	6.3%
MCO 3	239	13	13	5.4%	5.5%	5.4%	2.6%	8.3%
MCO 4	762	29	41	3.8%	5.4%	3.9%	2.3%	5.5%
MCO 5	5,139	275	281	5.4%	5.5%	5.4%	4.7%	6.0%
MCO 6	7,255	428	403	5.9%	5.6%	5.8%	5.3%	6.3%
MCO 7	2,962	168	158	5.7%	5.4%	5.8%	5.0%	6.6%
MCO 8	1,003	41	49	4.1%	4.9%	4.6%	3.1%	6.0%
MCO 9	352	18	17	5.1%	4.8%	5.8%	3.3%	8.4%
Total	28,711	1,572		5.5%				

Figure 17. Initiation and engagement of alcohol and other drug dependence treatment within 30 days of the initiation visit (IET-AD-30).

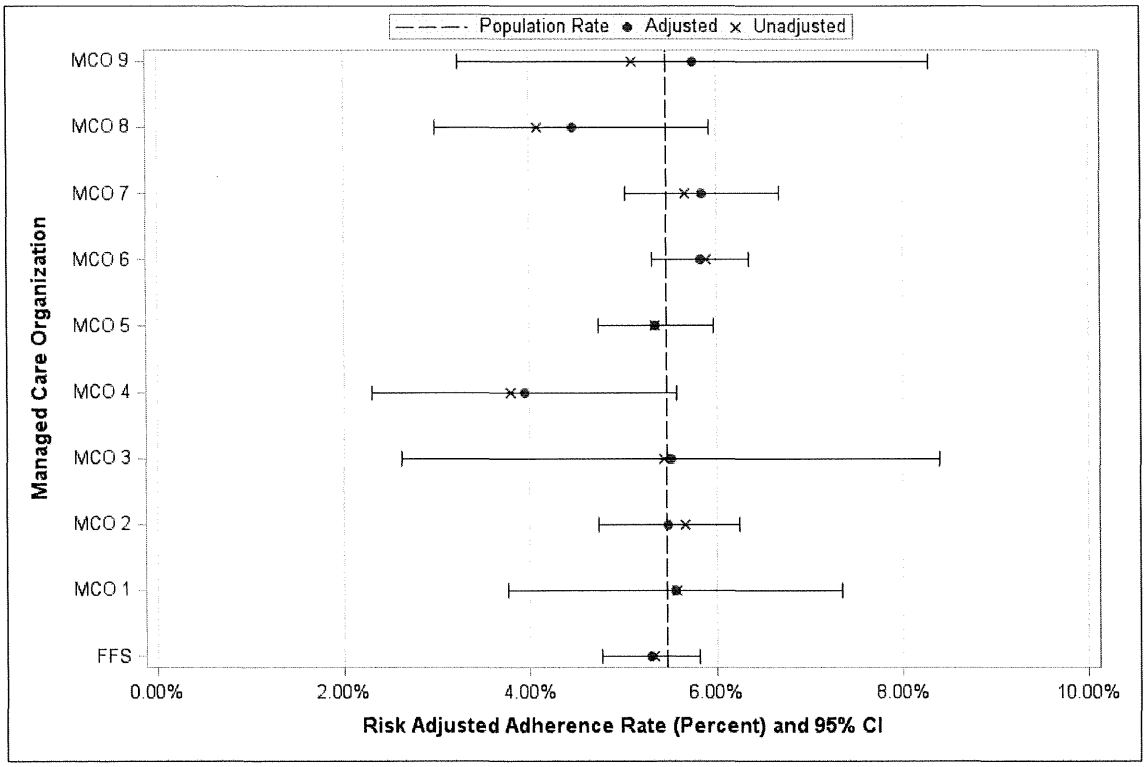


Table 54. Diabetes short-term complications admission rate (PQI01-AD).

MCO	Denominator (Member Months)	Observed (Members)	Expected (Members)	Observed Rate per 100,000 Member Months	Expected Rate per 100,000 Member Months	Risk Adjusted Rate per 100,000 Member Months	Lower 95% CI	Upper 95% CI
FFS	1,213,396	168	160	13.85	13.19	11.56	9.86	13.25
MCO 1	112,821	7	10	6.20	8.80	7.76	0.95	14.58
MCO 2	762,130	66	61	8.66	8.00	11.92	9.17	14.67
MCO 3	38,306	5	4	13.05	9.49	15.15	3.90	26.41
MCO 4	121,327	13	12	10.71	9.96	11.85	5.67	18.03
MCO 5	760,870	78	93	10.25	12.17	9.27	7.05	11.50
MCO 6	1,233,076	126	133	10.22	10.75	10.47	8.60	12.33
MCO 7	526,998	56	46	10.63	8.68	13.48	10.30	16.65
MCO 8	78,810	12	13	15.23	15.97	10.50	4.47	16.53
MCO 9	20,090	5	6	24.89	29.01	9.45	0.58	18.32
Total	4,867,824	536		11.01				

Figure 18. Diabetes short-term complications admission rate (PQI01-AD).

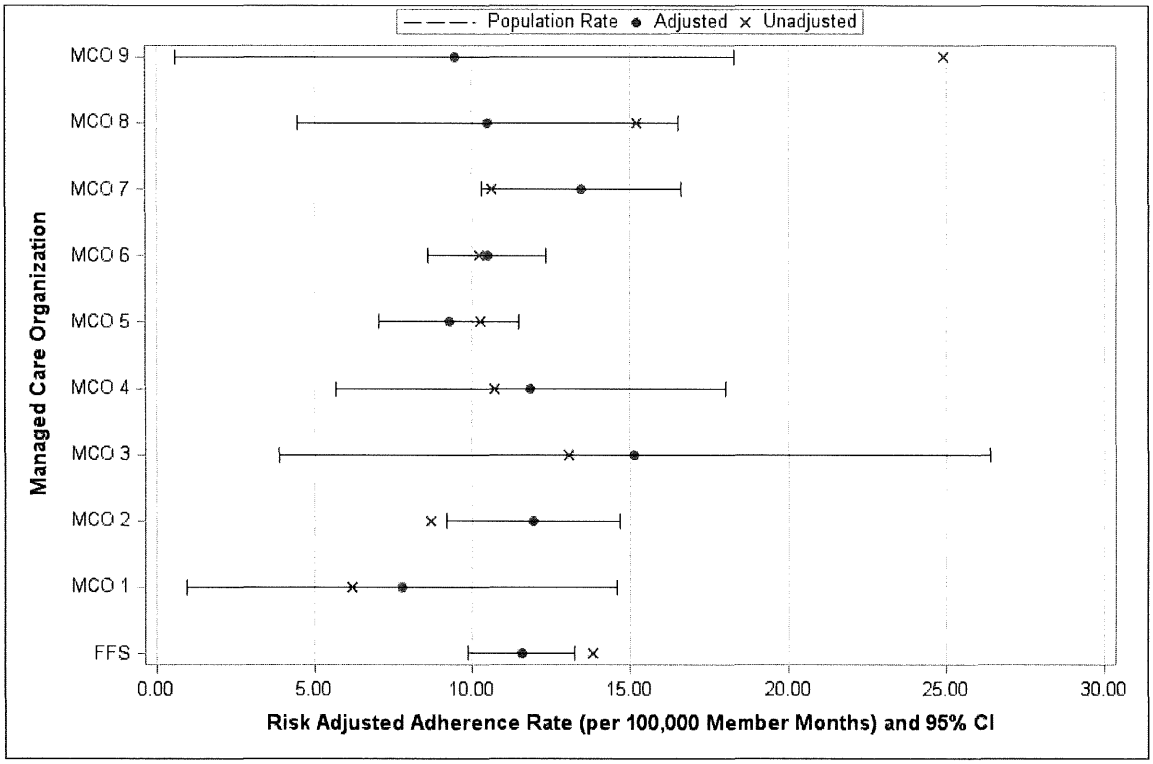


Table 55. Chronic obstructive pulmonary disease (COPD) or asthma in older adults admission rate (PQI05-AD).

MCO	Denominator (Member Months)	Observed (Members)	Expected (Members)	Observed Rate per 100,000 Member Months	Expected Rate per 100,000 Member Months	Risk Adjusted Rate per 100,000 Member Months	Lower 95% CI	Upper 95% CI
FFS	354,906	259	237	72.98	66.80	39.98	35.39	44.57
MCO 1	37,991	7	9	18.43	23.26	28.99	5.06	52.93
MCO 2	306,349	37	57	12.08	18.61	23.75	14.31	33.18
MCO 3	*	*	*	*	*	*	*	*
MCO 4	41,386	9	12	21.75	28.43	27.99	7.28	48.70
MCO 5	315,295	120	124	38.06	39.49	35.28	28.93	41.62
MCO 6	466,186	157	149	33.68	31.97	38.55	32.74	44.36
MCO 7	199,832	36	39	18.02	19.65	33.56	22.20	44.92
MCO 8	36,482	11	10	30.15	26.46	41.71	18.79	64.62
MCO 9	13,427	16	14	119.16	105.10	41.49	22.73	60.26
Total	1,786,992	654		36.60				

Note: * denotes values masked due to small numbers (n < 5).

Figure 19. Chronic obstructive pulmonary disease (COPD) or asthma in older adults admission rate (PQI05-AD).

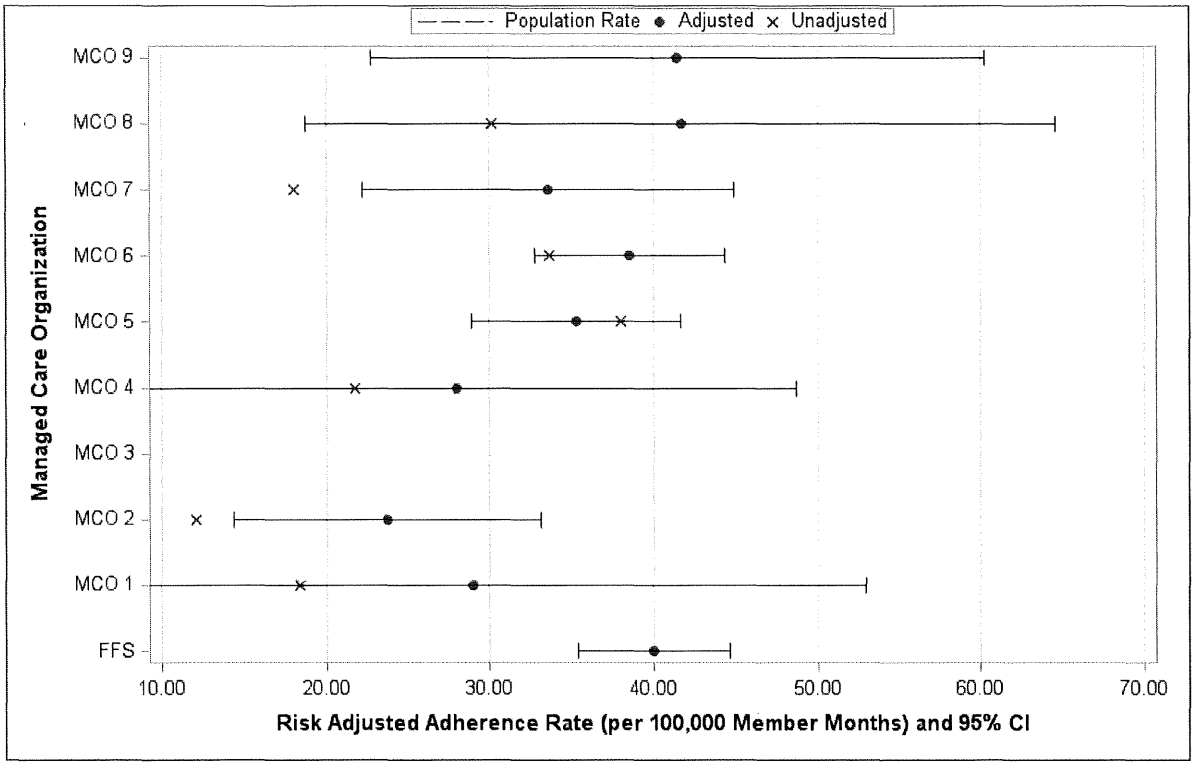
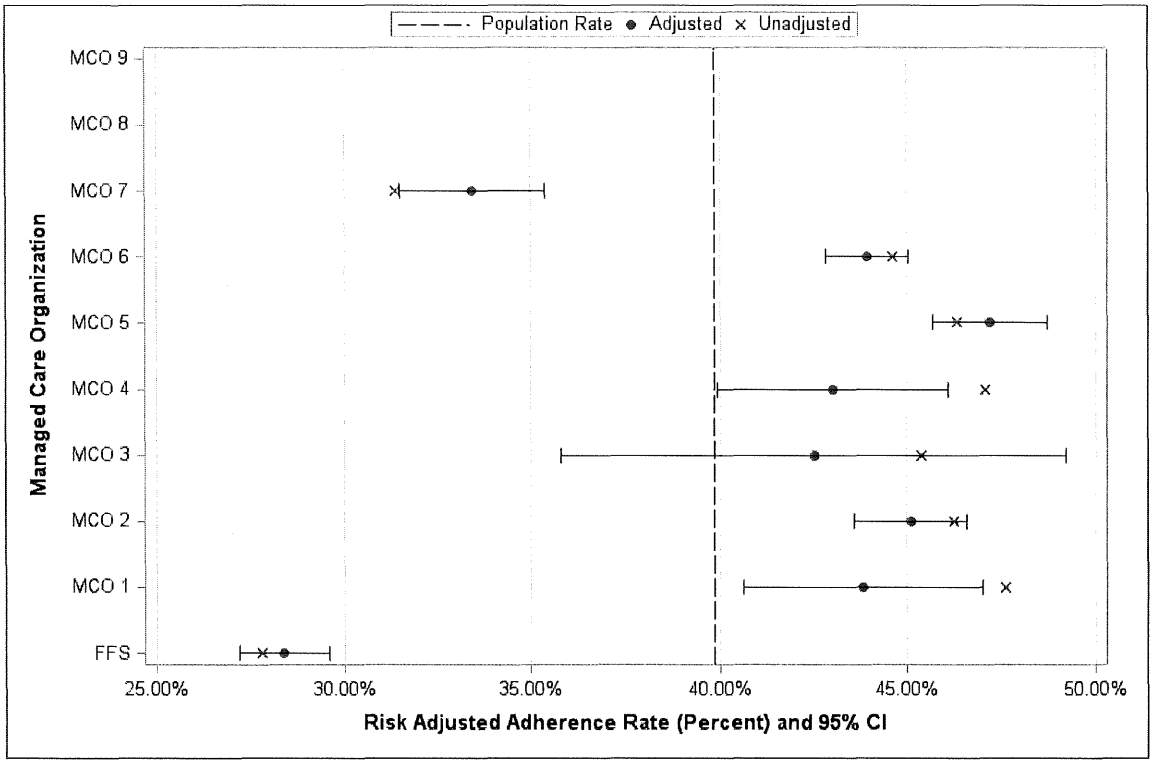


Table 56. Postpartum care rate (PPC-AD).

MCO	Denominator (Members)	Observed (Members)	Expected (Members)	Observed Rate	Expected Rate	Risk Adjusted Rate	Lower 95% CI	Upper 95% CI
FFS	6,331	1,761	2,470	27.8%	39.0%	28.4%	27.2%	29.6%
MCO 1	775	369	336	47.6%	43.3%	43.8%	40.6%	47.0%
MCO 2	3,849	1,779	1,572	46.2%	40.8%	45.1%	43.6%	46.6%
MCO 3	183	83	78	45.4%	42.5%	42.5%	35.8%	49.2%
MCO 4	824	388	359	47.1%	43.6%	43.0%	40.0%	46.1%
MCO 5	4,011	1,858	1,569	46.3%	39.1%	47.2%	45.7%	48.7%
MCO 6	7,405	3,302	2,995	44.6%	40.5%	43.9%	42.8%	45.0%
MCO 7	2,627	825	983	31.4%	37.4%	33.4%	31.5%	35.4%
MCO 8	*	*	*	*	*	*	*	*
MCO 9	*	*	*	*	*	*	*	*
Total	26,030	10,370		39.8%				

Note: * denotes values masked due to small numbers (n < 5).

Figure 20. Postpartum care rate (PPC-AD).



D. Stratified Risk Adjusted Rates

Table 57. Risk adjusted rates for the breast cancer screening measure (BCS-AD), stratified by age group and MCO.

MCO	Age	Observed Rate per 100,000 Member Months	Expected Rate per 100,000 Member Months	Risk Adjusted Rate per 100,000 Member Months	Lower 95% CI	Upper 95% CI
FFS	45 – 64	55.9%	57.3%	58.3%	56.9%	59.8%
FFS	65+	26.1%	36.1%	43.2%	32.1%	54.2%
MCO 1	45 – 64	62.3%	59.9%	62.1%	57.3%	67.0%
MCO 1	65+	28.6%	45.6%	37.5%	0.0%	84.9%
MCO 2	45 – 64	64.4%	63.3%	60.8%	59.4%	62.3%
MCO 2	65+	60.0%	49.1%	73.0%	56.4%	89.6%
MCO 3	45 – 64	68.2%	64.9%	62.7%	56.1%	69.4%
MCO 3	65+	100.0%	62.1%	96.2%	4.7%	100.0%
MCO 4	45 – 64	54.0%	62.1%	52.0%	47.8%	56.1%
MCO 4	65+	0.0%	51.6%	0.0%	0.0%	100.0%
MCO 5	45 – 64	59.3%	60.8%	58.3%	56.8%	59.7%
MCO 5	65+	51.8%	47.4%	65.3%	56.6%	74.0%
MCO 6	45 – 64	61.1%	60.2%	60.7%	59.4%	62.0%
MCO 6	65+	44.0%	47.6%	55.2%	48.9%	61.5%
MCO 7	45 – 64	65.9%	62.0%	63.5%	61.4%	65.7%
MCO 7	65+	65.9%	48.3%	81.4%	68.3%	94.5%
MCO 8	45 – 64	61.4%	65.3%	56.3%	49.6%	63.0%
MCO 8	65+	--	--	--	--	--
MCO 9	45 – 64	54.8%	55.8%	58.8%	52.0%	65.6%
MCO 9	65+	58.1%	49.0%	70.8%	49.6%	92.0%

Table 58. Risk adjusted rates for the breast cancer screening measure (BCS-AD), stratified by education and MCO.

MCO	Education	Observed Rate per 100,000 Member Months	Expected Rate per 100,000 Member Months	Risk Adjusted Rate per 100,000 Member Months	Lower 95% CI	Upper 95% CI
FFS	Unknown	57.9%	49.0%	70.5%	53.7%	87.4%
FFS	Less Than High School	50.8%	52.6%	57.7%	55.1%	60.4%
FFS	High School Graduate	57.1%	58.9%	57.9%	56.0%	59.9%
FFS	More than High School Graduate	56.9%	58.4%	58.2%	54.5%	61.9%
MCO 1	Unknown	80.0%	69.8%	68.5%	34.5%	100.0%
MCO 1	Less Than High School	57.6%	56.5%	61.0%	50.7%	71.3%
MCO 1	High School Graduate	61.5%	60.3%	61.0%	55.1%	66.9%
MCO 1	More than High School Graduate	70.3%	62.3%	67.4%	52.8%	82.1%
MCO 2	Unknown	59.5%	63.7%	55.8%	46.6%	65.0%
MCO 2	Less Than High School	66.1%	59.5%	66.3%	63.2%	69.5%
MCO 2	High School Graduate	63.9%	64.5%	59.2%	57.3%	61.1%
MCO 2	More than High School Graduate	64.2%	63.8%	60.2%	56.6%	63.8%
MCO 3	Unknown	66.7%	71.6%	55.7%	13.0%	98.3%
MCO 3	Less Than High School	62.2%	61.9%	60.0%	45.4%	74.7%
MCO 3	High School Graduate	69.4%	65.6%	63.2%	54.8%	71.6%
MCO 3	More than High School Graduate	75.0%	66.0%	67.9%	49.9%	85.9%
MCO 4	Unknown	72.7%	71.8%	60.5%	38.5%	82.6%
MCO 4	Less Than High School	53.0%	59.8%	53.0%	44.3%	61.6%
MCO 4	High School Graduate	54.0%	62.6%	51.5%	46.0%	57.0%
MCO 4	More than High School Graduate	52.2%	62.6%	49.8%	39.4%	60.2%
MCO 5	Unknown	52.6%	66.9%	47.0%	33.9%	60.1%
MCO 5	Less Than High School	54.9%	55.9%	58.7%	56.0%	61.4%
MCO 5	High School Graduate	60.7%	61.9%	58.6%	56.7%	60.5%
MCO 5	More than High School Graduate	62.1%	63.3%	58.6%	55.1%	62.2%
MCO 6	Unknown	71.1%	65.5%	64.8%	55.6%	74.1%
MCO 6	Less Than High School	52.6%	54.7%	57.5%	55.3%	59.6%
MCO 6	High School Graduate	65.5%	62.7%	62.4%	60.7%	64.2%
MCO 6	More than High School Graduate	65.4%	63.3%	61.7%	58.2%	65.2%
MCO 7	Unknown	71.0%	68.8%	61.7%	47.8%	75.5%

MCO	Education	Observed Rate per 100,000 Member Months	Expected Rate per 100,000 Member Months	Risk Adjusted Rate per 100,000 Member Months	Lower 95% CI	Upper 95% CI
MCO 7	Less Than High School	66.3%	57.7%	68.7%	64.7%	72.7%
MCO 7	High School Graduate	65.8%	62.9%	62.4%	59.5%	65.4%
MCO 7	More than High School Graduate	64.7%	63.4%	61.0%	55.6%	66.3%
MCO 8	Unknown	--	--	--	--	--
MCO 8	Less Than High School	60.0%	65.9%	54.4%	42.7%	66.1%
MCO 8	High School Graduate	60.5%	65.6%	55.1%	46.3%	63.9%
MCO 8	More than High School Graduate	70.6%	61.7%	68.3%	46.6%	90.0%
MCO 9	Unknown	--	--	--	--	--
MCO 9	Less Than High School	50.0%	52.8%	56.6%	47.4%	65.8%
MCO 9	High School Graduate	62.6%	57.4%	65.2%	55.2%	75.2%
MCO 9	More than High School Graduate	55.0%	58.0%	56.7%	34.5%	78.8%

Table 59. Risk adjusted rates for the breast cancer screening measure (BCS-AD), stratified by language and MCO.

MCO	Language	Observed Rate per 100,000 Member Months	Expected Rate per 100,000 Member Months	Risk Adjusted Rate per 100,000 Member Months	Lower 95% CI	Upper 95% CI
FFS	Non-English	45.0%	47.2%	57.0%	52.9%	61.0%
FFS	English	57.1%	58.6%	58.2%	56.6%	59.7%
MCO 1	Non-English	42.9%	48.5%	52.8%	30.9%	74.7%
MCO 1	English	63.2%	60.6%	62.4%	57.5%	67.3%
MCO 2	Non-English	69.1%	55.2%	74.8%	70.0%	79.6%
MCO 2	English	63.7%	64.2%	59.2%	57.7%	60.8%
MCO 3	Non-English	58.3%	56.5%	61.7%	32.5%	91.0%
MCO 3	English	69.2%	65.6%	63.0%	56.2%	69.9%
MCO 4	Non-English	66.7%	53.4%	74.6%	39.1%	100.0%
MCO 4	English	53.7%	62.3%	51.5%	47.3%	55.7%
MCO 5	Non-English	51.0%	51.3%	59.4%	55.3%	63.5%
MCO 5	English	60.6%	62.0%	58.4%	56.8%	59.9%
MCO 6	Non-English	47.8%	52.3%	54.6%	52.1%	57.1%
MCO 6	English	66.5%	63.1%	63.0%	61.5%	64.4%
MCO 7	Non-English	68.0%	54.1%	75.1%	69.8%	80.5%
MCO 7	English	65.3%	63.4%	61.6%	59.2%	63.9%
MCO 8	Non-English	69.2%	54.7%	75.7%	47.9%	100.0%
MCO 8	English	60.7%	66.2%	54.8%	47.9%	61.7%
MCO 9	Non-English	37.0%	47.5%	46.5%	32.4%	60.7%
MCO 9	English	62.4%	57.9%	64.4%	57.1%	71.7%

Table 60. Risk adjusted rates for the breast cancer screening measure (BCS-AD), stratified by metropolitan county and MCO.

MCO	Metropolitan County	Observed Rate per 100,000 Member Months	Expected Rate per 100,000 Member Months	Risk Adjusted Rate per 100,000 Member Months	Lower 95% CI	Upper 95% CI
FFS	No	55.3%	57.9%	57.1%	54.3%	60.0%
FFS	Yes	54.7%	56.1%	58.3%	56.6%	60.0%
MCO 1	No	62.3%	59.7%	62.3%	57.1%	67.5%
MCO 1	Yes	58.0%	59.3%	58.5%	45.6%	71.4%
MCO 2	No	65.3%	63.0%	61.9%	59.9%	63.9%
MCO 2	Yes	63.3%	63.2%	59.8%	57.6%	62.0%
MCO 3	No	68.4%	64.9%	63.0%	56.3%	69.6%
MCO 3	Yes	--	--	--	--	--
MCO 4	No	54.1%	62.1%	52.0%	47.8%	56.2%
MCO 4	Yes	33.3%	64.4%	30.9%	0.0%	80.6%
MCO 5	No	56.7%	60.7%	55.8%	52.0%	59.6%
MCO 5	Yes	59.3%	60.1%	58.9%	57.4%	60.5%
MCO 6	No	65.3%	60.9%	64.1%	60.6%	67.6%
MCO 6	Yes	59.2%	59.1%	59.9%	58.5%	61.2%
MCO 7	No	62.5%	53.2%	70.2%	34.4%	100.0%
MCO 7	Yes	65.9%	61.4%	64.2%	62.0%	66.3%
MCO 8	No	66.7%	71.9%	55.4%	13.1%	97.7%
MCO 8	Yes	61.3%	65.1%	56.3%	49.5%	63.1%
MCO 9	No	100.0%	60.6%	98.6%	31.8%	100.0%
MCO 9	Yes	54.9%	54.9%	59.7%	53.2%	66.2%

Table 61. Risk adjusted rates for the breast cancer screening measure (BCS-AD), stratified by race/ethnicity and MCO.

MCO	Race/ Ethnicity	Observed Rate per 100,000 Member Months	Expected Rate per 100,000 Member Months	Risk Adjusted Rate per 100,000 Member Months	Lower 95% CI	Upper 95% CI
FFS	Non-White	50.6%	53.0%	57.0%	54.6%	59.5%
FFS	White	57.6%	58.8%	58.6%	56.8%	60.4%
MCO 1	Non-White	57.5%	55.3%	62.1%	47.2%	77.0%
MCO 1	White	62.3%	60.3%	61.7%	56.7%	66.8%
MCO 2	Non-White	64.3%	58.8%	65.3%	60.7%	69.9%
MCO 2	White	64.4%	63.7%	60.4%	58.9%	62.0%
MCO 3	Non-White	61.5%	65.9%	55.8%	32.6%	79.0%
MCO 3	White	69.0%	64.8%	63.6%	56.6%	70.6%
MCO 4	Non-White	47.0%	56.7%	49.5%	37.4%	61.5%
MCO 4	White	55.1%	63.0%	52.2%	47.8%	56.7%
MCO 5	Non-White	56.9%	56.9%	59.7%	57.2%	62.2%
MCO 5	White	60.2%	62.1%	57.8%	56.1%	59.6%
MCO 6	Non-White	53.3%	55.2%	57.7%	55.7%	59.7%
MCO 6	White	66.0%	63.1%	62.5%	60.9%	64.2%
MCO 7	Non-White	69.0%	58.2%	70.7%	67.1%	74.4%
MCO 7	White	63.8%	63.4%	60.2%	57.5%	62.9%
MCO 8	Non-White	65.7%	65.4%	60.0%	51.8%	68.2%
MCO 8	White	52.9%	65.0%	48.7%	37.0%	60.3%
MCO 9	Non-White	55.9%	54.3%	61.5%	54.2%	68.7%
MCO 9	White	52.2%	58.0%	53.8%	39.2%	68.3%

Appendix D

The following notation was adapted from a document released by the Agency for Healthcare Research and Quality. [88]

Observed Rate

$$O_i = \left(\frac{1}{n_i}\right)\sum Y_{ij}$$

where the subscript j denotes a given member within a given group i (e.g., MCO)

Expected Rate

$$E_i = \left(\frac{1}{n_i}\right)\sum P_{ij}$$

Risk Adjusted Rate (RAR):

$$RAR_i = \frac{O_i}{E_i}$$

Variance of the RAR:

$$Variance = \left(\frac{\alpha}{E_i}\right)^2 \left(\frac{1}{n_i}\right)^2 \Sigma\left[P_{ij}(1 - P_{ij})\right]$$

Standard Error (SE):

$$SE = \sqrt{Variance}$$

95% Confidence Interval (CI) around RAR:

$$CI = RAR \pm (1.96 \times SE)$$

Equation Components:

Y_{ij} : 0 or 1 outcome (i.e., "yes" or "no") for a given member (j) in a given group (i)

P_{ij} : Predicted probability (i.e., score) of an outcome for a given member (j) in a given group (i)

n_i : unduplicated count of members in a given group (i)

α : rate in the entire population for a quality measure

Appendix E

The following section provides additional detail about how characteristics were derived and categorized.

Clinical Characteristics

Developmental Disability

- Source
  - Derived from EDC values from ACG data provided by DHS
- Definition
  - Yes
    - EDC Code(s)
      - NUR19: Developmental Disorder
  - No
    - None of these EDC codes.
- Comments
  - The assessment period date was restricted to EDC codes within the 2013 measurement year.
  - NUR abbreviation indicates Neurologic EDC measures.

Dialysis

- Source
  - Obtained from a dialysis service indicator variable from ACG data provided by DHS.
- Definition
  - N/A
- Comments
  - The frequency distribution of the dialysis service indicator variable was initially explored; no specific categories were created.
  - The dialysis measure was not included due to a lack of variation.
    - Only 0.09% were identified as having received dialysis while 99.91% were not identified as having received dialysis.

Disability

- Source
  - Derived from eligibility type from enrollment data provided by DHS
- Definition
  - Yes
    - 15: 1619A
    - 16: 1619B
    - BC: Breast / Cervical Cancer Control Program
    - BD: Blind / Prescription Drug
    - BQ: Blind / QMB Only
    - BS: Blind / SLMB
    - BT: Blind / TEFRA
    - BW: Blind / QWD
    - BX: Blind / No Sub-Type
    - DC: Disabled / Child Age 18 Thru 20
    - DI: Employed Disabled With No Premium
    - DP: MA For Employed Disabled With Premium
    - DQ: Disabled / QMB Only
    - DS: Disabled / SLMB
    - DT: Disabled / TEFRA

        - DW: Disabled / QWD
        - DX: Disabled / No Sub-Type
        - 1B: Qualifying Individual 1-Blind
        - 1D: Qualifying Individual 1 - Disabled
    - No
        - None of these eligibility type values

Frailty

- Source
    - Derived from the ACG frailty indicator variable provided by DHS
- Definition
    - Yes
        - Frailty indicator value was Y
    - No
        - Frailty indicator value was not Y

Mental Health

- Source
    - Derived from EDC values from ACG data provided by DHS
- Definition
    - Yes
        - EDC Code(s)
            - PSY01:Anxiety, neuroses
            - PSY04: Behavior problems
            - PSY05: Attention deficit disorder
            - PSY06: Family and social problems
            - PSY07: Schizophrenia and affective psychosis
            - PSY08: Personality disorders
            - PSY09: Depression
            - PSY10: Psychologic signs and symptoms
            - PSY11: Psychosocial disorders, other
            - PSY12: Bipolar disorder
    - No
        - None of these EDC codes.
- Comments
    - The assessment period date was restricted to EDC codes within the 2013 measurement year
    - PSY abbreviation indicates Mental Health / Psychosocial EDC measures.

Resource Utilization Band (RUB)

- Source
    - Obtained from the resource utilization bands variable in ACG data provided by DHS
- Definition
    - 0: Non-User
    - 1: Healthy
    - 2: Low Morbidity
    - 3: Moderate Morbidity
    - 4: High Morbidity
    - 5: Very High Morbidity

Substance Abuse

- Source
    - Derived from EDC values from ACG data provided by DHS
- Definition

    - Yes
        - EDC Code(s)
            - PSY02: Substance abuse
            - PSY03: Tobacco use
    - No
        - None of these EDC codes.
- Comments
    - The assessment period date was restricted to EDC codes within the 2013 measurement year
    - PSY abbreviation indicates Mental Health / Psychosocial EDC measures.

Sociodemographic Characteristics

Age

- Source
    - Derived from birthdate values provided by DHS enrollment data.
- Definition
    - Under 45 years
    - 45 to 64 years (i.e., less than 65 years)
    - 65 + years
- Comments
    - Age was calculated as of December 31, 2013.

Education

- Source
    - Derived from education values provided by DHS enrollment data.
- Definition
    - Unknown
        - Education value was missing (i.e., null)
    - Less than High School
        - Education value was not missing and was less than 12
    - High School Graduate
        - Education value was not missing and equal to 12
    - More than High School Graduate
        - Education value was not missing and greater than 12

Gender

- Source
    - Derived from sex values provided by DHS enrollment data
- Definition
    - Male
        - Sex value of M
    - Female
        - Sex value of F

Immigration Status

- Source
    - Obtained from immigration status values provided by DHS enrollment data
- Definition
    - N/A
- Comments
    - DHS decided not to include this characteristic due to a concern regarding the reliability of these values.

Language

- Source
  - Derived from language code values provided by DHS enrollment data.
- Definition
  - English
    - Language code value was 99
  - Non-English
    - Language code value was not 99

Metropolitan County

- Source
  - Derived from the member county of residence information provided by DHS enrollment data. The county information was then combined with rural-urban continuum codes (RUCC).[85]
- Definition
  - Metropolitan County
    - Yes
      - RUCC values
        - 1: Metro – Counties in metro areas of 1 million population or more
        - 2: Metro – Counties in metro areas of 250,000 to 1 million population
        - 3: Metro – Counties in metro areas of fewer than 250,000 population
    - No
      - None of these RUCC values.
- Comment:
  - The RUCC codes for each county are publicly available and downloadable on the USDA website.[85]

Race / Ethnicity

- Source
  - Derived from race and ethnicity indicator values provided by DHS enrollment data.
- Definition
  - White
    - Race value was W and Hispanic ethnicity indicator was not Y
  - Non-white
    - Race value was not W or Hispanic ethnicity indicator was Y

Families and Children Medical Assistance group

- Source
  - Derived from the ESPID, major program, eligibility type, and birthdate values provided by DHS enrollment data.
- Definition
  - Yes
    - MinnesotaCare: Children and adults who were neither seniors nor disabled.
      - ESPID Values
        - BB01: MinnesotaCare Basic Plus One
        - BB02: MinnesotaCare Limited Benefit
        - FF01: MinnesotaCare Basic Plus Two
        - FF02: MinnesotaCare Basic Plus
        - JJ01: MinnesotaCare Basic Plus Two
        - JJ02: MinnesotaCare Basic Plus

        - KK01: MinnesotaCare Expanded
        - LL01: MinnesotaCare Expanded
    - F+C MA: Families and Children Medical Assistance – Medicaid for children and adult parents and caretakers.
        - ESPID Values
            - GM01: General Assistance Medical Care
            - GM03: MinnesotaCare Basic Plus One
            - MA12: MA
            - MA13: Prepaid Medical Assistance Program (PMAP PIN)
            - MA20: MA
            - NM12: Medical Assistance < 65
    - FFS MA (non-disabled and non-senior): subset of Fee-for-service Medical Assistance who were not disabled and not seniors
        - Eligibility Type Values
            - Disability = No
                - Please see above description for creation of disability variable from specific eligibility type values.
        - Birthdate
            - Age < 65 years
                - Please see above description for creation of age.
        - ESPID Values
            - Null (i.e., blank or missing data)
        - Major Program Values
            - GM: General Assistance Medical Care
            - MA: Medicaid
            - NM: Non-Citizen Medical
- No
    - None of the above criteria
- Comments
    - This characteristic was not included due to collinearity with the disability variable.

<u>Utilization Characteristics</u>

Emergency Department Visit

- Source
    - Derived from an emergency department visit count variable from ACG data provided by DHS.
- Definition
    - Two or more visits
        - Visit count greater than one
    - One or fewer visits
        - Visit count less than two
- Comments
    - The decision was made not to include utilization variables in risk adjustment (see Utilization Characteristics section within Appendix A).

Inpatient Hospitalization

- Source
    - Derived from an inpatient hospitalization visit count variable from ACG data provided by DHS.
- Definition

    - Yes
        - Visit count greater than zero
    - No
        - Visit count not greater than zero
- Comments
    - The decision was made not to include utilization variables in risk adjustment (see Utilization Characteristics section within Appendix A).

Major Procedure

- Source
    - Obtained from a major procedure indicator variable from ACG data provided by DHS.
- Definition
    - N/A
- Comments
    - The frequency distribution of the procedure indicator variable was initially explored; no specific categories were created.
    - The decision was made not to include utilization variables in risk adjustment (see Utilization Characteristics section within Appendix A).

Nursing

- Source
    - Obtained from a nursing service indicator variable from ACG data provided by DHS.
- Definition
    - N/A
- Comments
    - The frequency distribution of the nursing service indicator variable was initially explored; no specific categories were created.
    - The decision was made not to include utilization variables in risk adjustment (see Utilization Characteristics section within Appendix A).

Outpatient Visit

- Source
    - Obtained from an outpatient visit count variable from ACG data provided by DHS.
- Definition
    - N/A
- Comments
    - The frequency distribution of the outpatient visit count was initially explored; no specific categories were created.
    - The decision was made not to include utilization variables in risk adjustment (see Utilization Characteristics section within Appendix A).

Providers (Count)

- Source
    - Derived from a unique provider count variable from ACG data provided by DHS.
- Definition
    - Five or more providers
    - Zero to five providers
- Comments
    - The decision was made not to include utilization variables in risk adjustment (see Utilization Characteristics section within Appendix A).

Saw a Specialist

- Source
    - Derived from a specialty count variable from ACG data provided by DHS.
- Definition

    - Yes
        - Count greater than zero
    - No
        - Count not greater than zero.
- Comments
    - The decision was made not to include utilization variables in risk adjustment (see Utilization Characteristics section within Appendix A).

Saw a Generalist

- Source
    - Derived from a generalist seen indicator variable from ACG data provided by DHS.
- Definition
    - Yes
        - Indicator value was Y
    - No
        - Indicator value was not Y
- Comments
    - The decision was made not to include utilization variables in risk adjustment (see Utilization Characteristics section within Appendix A).

References

[1] Kuhlthau K., Ferris TG., Iezzoni LI., 2004. Risk adjustment for pediatric quality indicators. *Pediatrics.* 113;210.

[2] Smith PC, Mossialos E, Papanicolas I, and Leatherman S., 2008. Performance measurement for health system improvement. *Cambridge University Press*. Available from: <http://www.euro.who.int/__data/assets/pdf_file/0009/135972/E94887_Part_III.pdf>

[3] Agency for HealthCare Research and Quality, 2010. Selecting quality and resource use measures: a decision guide for community quality collaboratives. Available from:< http://www.ahrq.gov/professionals/quality-patient-safety/quality-resources/tools/perfmeasguide/index.html>

[4] Scholle SH, Roski J, Adams JL, 2008. Benchmarking physician performance: reliability of individual and composite measures. *Am J Manag Care*, (12):833-8. Available from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2667340/>

[5] Casalino LP, Elster A, Eisenberg A, Lewis E, Montgomery J, Ramos D, 2007. Will pay-for-performance and quality reporting affect health care disparities? *Health* Aff , 26(3). Available from: <http://content.healthaffairs.org/content/26/3/w405.long>

[6] Birkmeyer JD, Kerr EA, Dimick JB, 2006. Improving the quality of quality measurement. Institute of Medicine Committee on Redesigning Health Insurance Performance Measures, Payment, and Performance Improvement Programs. Available from: <http://www.iom.edu/Reports/2005/Performance-Measurement-Accelerating-Improvement.aspx>

[7] National Quality Forum, 2014. Risk Adjustment & Sociodemographic Factors. Available from: < http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=75398>

[8] National Quality Forum. August 2014. Risk Adjustment for Socioeconomic Status or Other Sociodemographic Factors. Available from < http://www.qualityforum.org/Publications/2014/08/Risk_Adjustment_for_Socioeconomic_Status_or_Other_Sociodemographic_Factors.aspx>.

[9] Department of Veterans Affairs: Management Decision and Research Center; Washington, DC: VA Health Services Research and Development Service in collaboration with Association for Health Services Research, 1997. Risk adjustment: a tool for leveling the playing field. Available from: < http://www.hsrd.research.va.gov/publications/internal/riskadj.pdf>

[10] University of Maryland and Actuarial Research Corporation, 2003. A Guide to Implementing a Health-Based Risk-Adjusted Payment System for Medicaid Managed Care Programs. Available from: < http://www.hilltopinstitute.org/Publications/A%20Guide%20to%20Implementing%20a%20Health-Based%20Risk-Adjusted%20Payment%20System%20for%20Medicaid%20Managed%20Care%20Programs_March%202003.pdf>

[11] National Quality Forum, 2014. Risk adjustment and sociodemographic factors: draft recommendations and rationale. Available from: <http://www.qualityforum.org/Projects/n-r/Risk_Adjustment_and_Socioeconomic_Status/Draft_Recommendations_02102014.aspx>

[12] de Jager DJ, Grootendorst DC, Jager KJ, van Dijk PC, Tomas LMJ, Ansell D, Collart F, Finne P, Heaf JG, De Meester J, Wetzels JFM, Rosendaal FR, Dekker FW: Cardiovascular and non-cardiovascular mortality among patients initiating dialysis. JAMA 2009;302:1782–1789

[13] Nicholson JM, 2010. Risk adjustment in maternity care: the use of indirect standardization. University of Pennsylvania Health System. Available from: <http://www.dovepress.com/risk-adjustment-in-maternity-care-the-use-of-indirect-standardization-a5076>

[14] National Quality Measures Clearinghouse. Cost of care: total resource use population-based per member per month (PMPM) index. Agency for Healthcare Research and Quality. Available from: <http://www.qualitymeasures.ahrq.gov/content.aspx?id=38364>

[15] Center for Medicaid and CHIP Services, 2013. Initial core set of health care quality measures for adults enrolled in Medicaid (Medicaid Adult Core Set): technical specifications and resource manual for federal fiscal year 2013. *Centers for Medicare and Medicaid Services.* Available from: <www.medicaid.gov/Medicaid.../Medicaid-Adult-Core-Set-Manual.pdf>

[16]Gilmer T. Risk adjustment using CDPS. Research Data Assistance Center. Available from: <http://www.resdac.org/sites/resdac.org/files/Risk%20Adjustment%20Using%20CDPS%20%28Slides%29.pdf>

[17] Shaughnessy PW and Hittle DF, 2002. Overview of the Risk adjustment and Outcome Measures for Home Health Agency OBQI Reports: Highlights of Current Approaches and Outline of Planned Enhancements. Center for Health Services Research, University of Colorado Health Sciences Center. Available from: <http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HomeHealthQualityInits/downloads/HHQIRiskAdj.pdf>

[18] Shwartz M, Ash A, Pekoz E, 2006. Risk adjustment and risk-adjusted provider profiles. *Int. J. Healthcare Technology and Management,* (7):1/2.

[19] Centers for Medicare and Medicaid Services, 2010. Outcome-Based Quality Improvement (OBQI) Manual. Available from: <http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HomeHealthQualityInits/downloads/HHQIOBQIManual.pdf>

[20] Duncan IG, 2011. Healthcare Risk Adjustment and Predictive Modeling. *Actex Publications.*

[21] Winkleman R, Mehmud S, 2007. A Comparative Analysis of Claims-Based Tools for Health Risk Assessment. Society of Actuaries. Available from: <http://www.soa.org/research/research-projects/health/hlth-risk-assement.aspx>

[22] Ingenix, 2014. Symmetry Pharmacy Risk Groups. Available from: <http://www.optuminsight.com/content/attachments/SymmetryPharmacyRiskGroupsproductsheet.pdf>

[23] Kronick R, Gilmore T, Dreyfus T, Lee L, 2000. Improving Health-Based Payment for Medicaid Beneficiaries: CDPS. University of California, San Diego. Available from: <http://cdps.ucsd.edu/cdps_hcfr.pdf>

[24] Stukenborg GJ, Wagner DP, Harrell FE, Jr., et al., 2007. Which hospitals have significantly better or worse than expected mortality rates for acute myocardial infarction patients? Improved risk adjustment with present-at-admission diagnoses. *Circulation*,116(25):2960-8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18071076>

[25] Fiscella K, Franks P, Gold MR, et al, 2000. Inequality in quality: addressing socioeconomic, racial, and ethnic disparities in health care. *JAMA*, 283(19):2579-84. Available from:<http://www.ncbi.nlm.nih.gov/pubmed/10815125>

[26] Romano P, Hussey P, Ritley D, 2010. Selecting quality and resource use measures: a decision guide for community quality collaboratives. *Agency for Healthcare Research and Quality*, 09(10)-0073. Available from: <http://www.ahrq.gov/professionals/quality-patient-safety/quality-resources/tools/perfmeasguide/index.html>

[27] Grantee-to-Grantee Discussion on Plan All Cause Readmission measure, Centers for Medicare and Medicaid *Services* (February 20, 2014).

[28] National Quality Measures Clearinghouse. Available from: <http://www.qualitymeasures.ahrq.gov/content.aspx?id=47277>

[29] Centers for Medicare and Medicaid Services, 2013. Specifications for the All-Cause Unplanned Readmission Measure for 30 Days Post Discharge from Long-Term Care Hospitals. Available from: <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/LTCH-Quality-Reporting/Downloads/LTCH-Readmissions-Measure-Specifications.pdf >

[30] Centers for Medicare and Medicaid Services, 2013. Specifications for the All-Cause Unplanned Readmission Measure for 30 Days Post Discharge from Inpatient Rehabilitation Facilities. Available from: <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/IRF-Quality-Reporting/Downloads/DRAFT-Specifications-for-the-Proposed-All-Cause-Unplanned-30-day-Post-IRF-Discharge-Readmission-Measure.pdf >

[31] Minnesota Department of Health, 2011. Minnesota Statewide Quality Reporting and Measurement System: Risk Adjustment of Physician Clinic Quality Measures. Available from: <http://www.health.state.mn.us/healthreform/measurement/MDHRiskAdjustmentProposal.pdf>

[32] Minnesota Community Measurement. 2013. 2013 Health care quality report. Available from: <http://mncm.org/wp-content/uploads/2014/02/2013-HCQR-Final-2.4.2014.pdf>

[33] McMahon LF, Hofer TP, Hayward RA, 2007. Physician-level P4P--DOA? Can quality-based payment be resuscitated? *Am J Manag Care*. 13(5):233-6. Available from:< http://www.ncbi.nlm.nih.gov/pubmed/17488186>

[34] Sarquis AL, Miyaki M, Cat MN, 2002. The use of CRIB score for predicting neonatal mortality risk. *J Pediatr (Rio J).* (3):225-9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/14647779>

[35] Vad den Akker M, Buntinx F., Metsmakers JF, Roos S, Knottnerus JA., 1998. Multimorbidity in general practice: prevalence, incidence, and determinants of co-occurring chronic conditions and recurrent diseases. *J Clin Epidemiol.* 51:367-375. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/9619963>

[36] Kuhlthau KA, Ferris TG, Beal AC, Gortmaker SL, Perrin JM, 2001. Who care for Medicaid-enrolled children with chronic conditions? *Pediatrics.* 108:906-912. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11581443>

[37] Forrest CB, Reid RJ, 2001. Prevalence of health problems and primary care physicians' specialty referral decisions. *J Fam Pract*, 50:427-432.

[38] Perrin JM, Kuhlthau K, Gortmaker SL, Beal AC, Ferris TG, 2003. Generalist and subspecialist care for children with chronic conditions. *Ambul Pediatr.*

[39] National Committee for Quality Assurance, 2010. Insights for Improvement: Measuring Health Care Value: Relative Resource Use. Available from: <http://www.ncqa.org/portals/0/hedisqm/RRU/BI%20NCQA_RRU_Publication_FINAL.pdf>

[40] National Committee for Quality assurance, 2014. HEDIS and Quality Compass. Available from: <http://www.ncqa.org/HEDISQualityMeasurement/WhatisHEDIS.aspx>

[41]Quality Compass. Interpreting the Measures. *National Committee for Quality Assurance.* Available from: <http://www.qualitycompass.org/QcsExternal/docs/InterpretingtheMeasures.pdf>

[42] National Committee for Quality Assurance. An Overview of NCQA Relative Resource Use Measures. Availble from: <http://www.ncqa.org/portals/0/hedisqm/RRU/BI%20NCQA_RRU_Publication_FINAL.pdf>

[43] Zaslavsky AM, Zaborski LB, Ding L, Shaul JA, Cioffi MJ, Cleary PD, 2001. Adjusting performance measures to ensure equitable plan comparisons. *Health Care Financ Rev.* 22:109-126. Available from: < http://www.hcp.med.harvard.edu/publications/adjusting-performance-measures-to-ensure-equitable-plan-comparisons>

[44] Mukamel DB, Brower CA. The Influence of Risk Adjustment Methods on Conclusions About Quality of Care in Nursing Homes Based on Outcome Measures. *The Gerontologist*, 38(6):695-703. Available from: <http://gerontologist.oxfordjournals.org/content/38/6/695.abstract>

[45] Berlowitz DR, Ash AS, Brandeis GH, Brand HK, Halpern JL, Moskowitz MA, 1996. Rating long-term care facilities on pressure ulcer development: importance of case-mix index. *Annals of Internal Medicine.* 124(6):557-563. Available from: < http://annals.org/article.aspx?articleid=709520>

[46] Rosen A, Wu J, Chang BH, et al., 2001. Risk adjustment of measuring health outcomes: an application in VA long-term Care. *Am J Med Qual.* 16(4):1180127.

[47] Hannan EL, Kilburn H, Racz M, Shields E, Chassin MR, 1994. Improving the outcomes of coronary artery bypass surgery in New York State. *JAMA,* 58:1852-7.

[48] Mukamel DB, Mushlin AI, 1998. Quality of care information makes a difference: an analysis of market share and price changes after publication of the New York State Cardiac Surgery Mortality Reports. *Med Care*, 36:945-54. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/9674613>

[49] Office of Statewide Health Planning and Development. Risk Adjustment for CABG Surgery Outcomes. Available from: < http://www.oshpd.ca.gov/HID/CORC/TrainingVideos/RiskAdjustmentSlides.pdf>

[50] Li Z, Carlisle DM, Marcin JP, Castellanos LR, Romano PS., Young JN, Amsterdam EA, 2010. Impact of Public Reporting on Access to coronary Artery Bypass Surgery: The California Outcomes Reporting Program. *Ann Thorac Surg*, 89(4):1131-8. Available from: < http://www.annalsthoracicsurgery.org/article/S0003-4975%2810%2900137-2/abstract>

[51]Office of Statewide Health Planning and Development, 2006. The California Report on Coronary Artery Bypass Graft Surgery: Hospital Responses. Available from: <.http://www.oshpd.ca.gov/HID/Products/PatDischargeData/CABG/03-04hospresponses.pdf >

[52] Raudenbush SW, Bryk AS, 2002. Hierarchical Linear Models: Applications and Data Analysis Methods. Second Edition. *Sage Publications.*

[53] Thomas N, Longford NT, Rolph JE, 1994. Empirical Bayes methods for estimating hospital-specific mortality rates. *Stat Med.* 13:889-903.

[54] Agency for Healthcare Research and Quality. Inpatient Quality Indicators. Available from: <http://www.qualityindicators.ahrq.gov/Downloads/Modules/IQI/V42/Inpatient_Broch_10_Update.pdf>

[55] Agency for Healthcare Research and Qualtiy. AHRQ Quality Indicators Risk Adjustment and Hierarchical Modeling Approaches (Draft Report).

[56] Dimick JB, Osborne NH, Hall BL, Ko CY, Birkmeyer JD, 2010. Risk adjustment for comparing hospital quality with surgery: how many variables are needed? J Am Coll Surg. 210(4):503-8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20347744>

[57] Tu JV, Sykora K, Naylor CD, 1997. Assessing the outcomes of coronary artery bypass graft surgery: how many risk factors are enough? Steering Committee of the Cardiac Care network of Ontario. *J Am Coll Cardiol;* 30:1317-1323. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/9350934>

[58] Minnesota Community Measurement. 2012. 2012 Health Care Disparities Report for Minnesota Health Care Programs. Available from: <http://mncm.org/wp-content/uploads/2013/04/Final2012HealthCareDisparitiesReport5.10.pdf>

[59] Inkelas M, Decristofaro AH, McGlynn EA, Keeler EB, 2000. Outcome Measurement in HEDIS: Can Risk Adjustment Save the Low Birth Weight Measure? *Health Serv Res,* 35(5 Pt 3):72-85. Available from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1383596/>

[60] Romano PS, Chan BK, Schembri ME, Rainwater, JA. 2002. Can administrative data be used to compare postoperative complication rates across hospitals? *Medical* Care, 40(10):856-867. Available from: <

http://www.jstor.org/discover/10.2307/3767575?uid=3739936&uid=2129&uid=2&uid=70&uid=4&uid=3739256&sid=21103874945043>

[61] Centers for Disease Control and Prevention, 2014. Classification of Diseases, Functioning, and Disability. Available from: <http://www.cdc.gov/nchs/icd/icf.htm >

[62] Bleich SN, Ozaltin Em, Murray CJ, 2009. How does satisfaction with the health-care system relate to patient experience? World Health Organization Bull. *World Health Organ* 2009;87:271-278. Available from: <http://www.who.int/bulletin/volumes/87/4/07-050401.pdf >

[63] Cumming RB, Cameron BA, 2002. A Comparative Analysis of Claims-based Methods of Health Risk Assessment for Commercial Populations. Society of Actuaries. Available from: <www.soa.org/Files/Research/_asset_id=2583046.pdf>

[64] Cummings RB, Knutson D, Cameron BA, Derrick B, 2002. A Comparative Analysis of Claims-based Methods of Health Risk Assessment for Commercial Populations. Society of Actuaries. Available from: <www.soa.org/sections/riskadjusfinalreport1.pdf>

[65] Shwartz M, Ash A, Pekoz E, 2006. Risk adjustment and risk-adjusted provider profiles. *Int. J. Healthcare Technology and Management*. (7):1/2.

[66] Jennings BM. Chapter 23 Patient Acuity. Patient Safety and Quality: An Evidence-Based Handbook for Nurses: Vol. 2.

[67] Preyra C, 2004. Coding Response to a Case-Mix Measurement System Based on Multiple Diagnosis. *Health Serv Res*, 39(4 Pt 1): 1027-1046. Available from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1361050/>

[68] Romano P, Hussey P, Ritley D, 2010. Selecting quality and resource use measures: a decision guide for community quality collaboratives. Agency for Healthcare Research and Quality, 09(10)-0073. Available from: <http://www.ahrq.gov/professionals/quality-patient-safety/quality-resources/tools/perfmeasguide/index.html>

[69] Rosen AK, Reid R, Broemeling A-M, Rakovski CC, 2003. Applying a risk-adjusted framework to primary care: can we improve on existing measures? *Annals of Family Medicine*, 1(1):44-51. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1466561/>

[70] Miller R, West C, 2007. The value of electronic health records in community health centers: policy implications. *Health Affairs,* 26(1):206-214. Available from: <http://content.healthaffairs.org/content/26/1/206.full>

[71] Casalino L, Elster A. 2007. Will pay-for-performance and quality reporting affect health disparities? *Health Affairs,* 26(3): 405-414. Available from: < http://content.healthaffairs.org/content/26/3/w405.long>

[72] Persell S, Wright J, Thompson J, Kmetik K, Baker D, 2006. Assessing the validity of national quality measures for coronary artery disease using an electronic health record. *Archives of Internal Medicine*, 166(20):2272-2277.

[73] Scott I, Youlden D, Coory M, 2004. Are diagnosis specific outcome indicators based on administrative data useful in assessing quality of hospital care? *Qual Saf Health Care,* 13(1):32-39. Available from: < http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1758063/>

[74] Krumholz HM, Wang Y, Mattera JA, et al., 2006. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with an acute myocardial infarction. *Circulation,* 113(13):1683-92. Available from: < http://www.ncbi.nlm.nih.gov/pubmed/16549637>

[75] Scholle SH, Roski J, Adams JL, et al., 2008. Benchmarking physician performance: reliability of individual and composite measures. *Am J Manag Care*, (12):833-8. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2667340/>

[76] Greenfield S, Kaplan SH, Kahn R, Ninomiya J, Griffith JL., 2002. Profiling care provided by different groups of physicians: effects of patient case-mix (bias) and physician-level clustering on quality assessmen results. *Ann Intern Med*, 136(2):111. Available from: <http://annals.org/article.aspx?articleid=668674>

[77] Ghosh D, Vogt A, 2012. Outliers: An evaluation of methodologies. Section on Survey Research Methods—JSM. American Statistical Association. Available from: <http://www.amstat.org/sections/srms/proceedings/y2012/files/304068_72402.pdf>

[78] Cary LA, Perou CM, Livasy CA, Dressler LG, Cowan K, Karaca G, et al., 2006. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *Jama.* (21):2492-502. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16757721>

[79] The Johns Hopkins ACG® System. <http://acg.jhsph.org/>

[80] The Johns Hopkins ACG® System. (2009). Technical Reference Guide Version 9.0. The Johns Hopkins University.

[81] Symmetry® EBM Connect® <https://www.optum.com/government/state/analytics-hit/population-analytics/measuring-value-understanding-costs/symmetry-ebm-connect.html>

[82] National Quality Forum. August 2014. Risk Adjustment for Socioeconomic Status or Other Sociodemographic Factors. Available from < http://www.qualityforum.org/Publications/2014/08/Risk_Adjustment_for_Socioeconomic_Status_or_Other_Sociodemographic_Factors.aspx>.

[83] Center for Medicaid and CHIP Services. Centers for Medicare & Medicaid Services. April 2015. Core Set of Adult Health Care Quality Measures for Medicaid (Adult Core Set). Technical Specifications and Resource Manual for Federal Fiscal Year 2015 Reporting. Available from: <http://www.medicaid.gov/Medicaid-CHIP-Program-Information/By-Topics/Quality-of-Care/Downloads/Medicaid-Adult-Core-Set-Manual.pdf>.

[84] The Mayo Clinic. Tests and Procedures – Mastectomy Definition. Available from: <http://www.mayoclinic.org/tests-procedures/mastectomy/basics/definition/prc-20012749>.

[85] United States Department of Agriculture. (2013). Economic Research Service. Available from: < http://www.ers.usda.gov/data-products/rural-urban-continuum-codes.aspx>.

[86] National Quality Forum. July 2014. Risk Adjustment for Socioeconomic Status or Other Sociodemographic Factors.

[87] Kleinbaum KM, Kupper LL, Muller KE, Nizam A. (1998). Applied regression analysis and other multivariable methods. Pacific Grove, CA: Duxbury Press.

[88] Agency for Healthcare Research and Quality. (2011). Calculating Standard Errors and Confidence Intervals for the AHRQ Quality Indicators. Available from < http://www.qualityindicators.ahrq.gov/Downloads/Resources/Publications/2011/Calculating_Confidence_Intervals_for_the_AHRQ_QI.pdf>.