MILLIMAN REPORT

Applying an equity lens to Medicaid risk adjustment

December 2024

Anders Larson, FSA, MAAA
Erica Baird, FSA, MAAA, PhD
Nazanin Dameshghi
Libby Bunzli, MPH
Spencer Polak, ASA, MAAA

Table of contents

Introduction

BACKGROUND

There is a growing desire at the federal and state levels of government to address health disparities that have proven to be persistent over time. State Medicaid agencies, as purchasers of healthcare services for diverse populations, are uniquely positioned to use their financing and policy tools thoughtfully and proactively to address health inequities.

A growing body of research continues to illuminate the impact of health-related social needs on health outcomes and healthcare costs. As such, the application of these insights into Medicaid financing mechanisms as an approach to improve health equity is burgeoning. Specifically, states have begun to explore the inclusion of social risk factors in risk adjustment methodologies as an enhancement to traditional clinical and diagnostic predictors of healthcare costs. Because individuals in certain demographic groups can face unique barriers to access and may feel cultural mistrust toward the healthcare system, it is important that risk adjustment models are developed thoughtfully to avoid inadvertently underinvesting in communities in need and exacerbating racial and ethnic inequities.

To help state Medicaid agencies and other stakeholders make informed decisions about strategies to address health equity, we undertook a rigorous research effort to examine the relationship between various factors (clinical, demographic, racial/ethnic, geographic, social) and healthcare utilization patterns. This research was designed to be applicable to a Medicaid managed care environment, in which payments to managed care organizations (MCO) typically vary based on factors such as age, gender, disability, and health status (as measured by risk scores). Based on findings from the analytical portion of our research, we explored policy options that can support states in applying more equitable risk adjustment and developing complementary activities to direct more healthcare resources and supportive programming into communities with higher social needs and health inequities.

STRUCTURE OF REPORT

This paper begins with a discussion of risk adjustment in Medicaid to give the reader a solid understanding of how risk score algorithms are currently used to allocate Medicaid funds. We also provide a comparison of two prominent risk score algorithms: Chronic Illness and Disability Payment System (CDPS) and Milliman Advanced Risk Adjusters (MARA). This comparison is intended to help readers understand the predictive accuracy of these two risk adjustment algorithms, which helps frame the amount of variation that may be explained by other factors. We conclude this section with commentary on how risk adjusters can be customized or otherwise enhanced for use in specific applications.

Next, we provide a detailed description of our analysis of healthcare utilization patterns using data from two state Medicaid agencies. After thoroughly discussing the data sources, methodology, and key assumptions, we recap the key findings of our analysis, with quantitative support and commentary about the implications of each finding.

We conclude with a discussion of potential policy options related to advancing health equity and addressing health-related social needs. We deliberately began work on this section after completing the analysis so that we focused on policy actions that specifically address disparities supported by our analysis. This section goes beyond a review of approaches used by states today; rather, we include commentary about the policies we believe would be most effective, leveraging insights gained from our work with state Medicaid agencies over the past three decades.

CAVEATS AND AREAS FOR FURTHER STUDY

Drivers of healthcare utilization are inherently complex and individual health outcomes and service utilization cannot be explained solely by predictive models. The analysis in this paper required difficult decisions regarding the modeling approach, which includes implicit assumptions about the relationships between the selected variables and the outcomes of interest. Although we believe the high-level conclusions generated by the results of our predictive model are reasonable and provide useful insights into the variables of interest, it is certain that different modeling choices may have resulted in different interpretations of the data.

Our analysis fit a single predictive model to all enrollees across both states. Although we reflected a wide range of features, the model did not include interaction terms or develop separate coefficients for different subpopulations. Future studies may consider diving deeper into specific subpopulations; for instance, fitting models specifically for pediatric or disabled populations.

Although the sample size of our analysis was large (more than 3 million enrollees), it was still limited to only two state Medicaid programs. There exists wide variation among state Medicaid programs across the country, including benefit packages, eligibility criteria, reimbursement structures, managed care dynamics, socioeconomic conditions, and mix of enrollees. Findings of this analysis will not necessarily generalize to all Medicaid programs or to other insurance markets.

The policy options presented in this report are not exhaustive and may not be appropriate for any particular Medicaid program. We encourage policymakers and other stakeholders to consider how the dynamics of their state affect the feasibility of the options we describe.

Milliman has developed certain models to estimate the values included in this report. The intent of the models was to analyze healthcare utilization patterns. We have reviewed the models, including their inputs, calculations, and outputs, for consistency, reasonableness, and appropriateness to the intended purpose and in compliance with generally accepted actuarial practice and relevant actuarial standards of practice (ASOP). The models rely on data and information as input to the models. We have relied upon certain data and information provided by state Medicaid agencies and other public data for this purpose and accepted it without audit. To the extent that the data and information provided is not accurate, or is not complete, the values provided in this report may likewise be inaccurate or incomplete.

Milliman's data and information reliance includes eligibility data, Medicaid fee-for-service (FFS) claims and managed care encounter data, the National Provider Identifier (NPI) registry, and publicly available geographic indicators. The models, including all input, calculations, and output, may not be appropriate for any other purpose.

It should be emphasized that many of the findings in this report implicitly reflect projections of future utilization based on a set of assumptions. Results will differ if actual experience is different from the assumptions contained in this analysis.

The materials in this document represent the opinions of the authors and do not necessarily represent the views of Milliman.

LITERATURE REVIEW

We conducted a review of academic and grey literature to understand progress in incorporating indicators of social need in risk adjustment. We found several studies that evaluate the impact of incorporating social risk factors into risk adjustment, either as geographic indices using publicly available data or as individual-level risk factors. The majority of published research explored the way these factors could affect the predictive accuracy of a risk score algorithm. Additionally, there were other studies that have analyzed disparities in certain types of utilization between different cohorts.

We believe our research approach differs from existing literature on these topics in two key ways:

1. We incorporated a broad range of features into the modeling process, with the goal of identifying the strongest relationships. Similarly, we built predictive models for multiple utilization types (such as emergency department visits and primary care visits) to gain a comprehensive understanding of the utilization patterns and how they vary based on the different patient characteristics.

2. The policy options considered to address the disparities identified in our analysis extend beyond incorporating additional features into the risk score algorithm. We explore policy approaches that can be complemented with adjustments to managed care payments to combine financial incentives with other program requirements to advance a multifaceted health equity strategy.

Throughout this paper, we have included references to other literature that support the findings from our research, provide additional context, or otherwise allow readers to deepen their understanding of specific topics.

Summary of key findings

Our research makes clear that, although clinical risk adjustment algorithms have strong predictive accuracy, even with minimal customization, significant variation remains in utilization patterns among different cohorts, which is not accounted for through risk scores. This variation can be partially explained by other factors that are not typically included in managed Medicaid risk adjustment mechanisms. In some cases, the variation suggests unequal access to quality healthcare and/or differences in care-seeking behavior, possibly driven by mistrust of the healthcare system by individuals within certain groups. However, addressing these disparities through Medicaid policy is more complex than simply incorporating these factors into risk adjustment algorithms.

KEY FINDINGS FROM ANALYSIS OF HEALTHCARE UTILIZATION PATTERNS

We analyzed claims, enrollment, and demographic data from July 2020 through June 2022 for more than 3 million Medicaid enrollees in two states to understand the relationship between clinical, demographic, socioeconomic, and geographic factors and various healthcare utilization measures. Although clinical risk scores were strongly predictive of many key utilization measures, we were most interested in the incremental impact of other factors that are often not accounted for in the risk adjustment process for managed Medicaid programs.

The factors that were most strongly associated with variation in utilization included:

- **Race/ethnicity.** Black individuals tended to have relatively lower primary care and retail pharmacy utilization and higher inpatient hospital and emergency department utilization compared to members of other race/ethnicity cohorts. Asian and Hispanic individuals tended to have relatively low utilization of most services, with the exception of primary care visits for Hispanic individuals, which were particularly high. White individuals had relatively average utilization of most services, although they had substantially higher utilization of antidepressants and antianxiety agents.
- **Homelessness.** People experiencing homelessness had much higher utilization of inpatient and emergency department services than other similarly situated individuals. Conversely, people in this cohort tended to have lower utilization of other services we reviewed, with the exception of antidepressants. These findings suggest a lack of ambulatory care to manage conditions, which may exacerbate the need for acute services.
- **Rural/urban status.** Individuals in rural areas had lower utilization of all services in our analysis, with the exception of inpatient admissions. This could be due to challenges accessing ambulatory care due to long driving distances and/or fewer providers per capita.
- **Problems related to upbringing.[1]** Individuals who had been identified through diagnosis codes as having problems related to their upbringing (such as being in custody of a non-parental relative or a history of parental abuse) had generally higher utilization of all services, particularly antidepressants and antianxiety agents. Other than homelessness, these were the strongest relationships we identified among "Z-codes," which are ICD-10 diagnosis codes used to identify social determinants of health.

Throughout this paper, we will comment on how utilization varied based on different factors. Unless otherwise noted, the statements refer to the effect of specific factors as measured by the regression analysis. The observations will differ from the results of a straightforward summary of the data for certain cohorts. For instance, when we state that individuals in rural areas had higher utilization of inpatient services, this means that, for two individuals with similar characteristics, the individual in the rural area is expected to have higher utilization than the individual in the urban area. It is possible that the prevalence of other factors (race/ethnicity, homelessness, age/gender mix, etc.) in rural versus urban areas could cause the overall utilization in these areas to differ from the findings of the regression analysis.

1. CMS. Improving the Collection of Social Determinants of Health (SDOH) Data with ICD-10-CM Z Codes. Retrieved November 26, 2024, from https://www.cms.gov/files/document/cms-2023-omh-z-code-resource.pdf.

These findings are discussed in more detail later in this report with additional graphical and numerical support, as well as commentary about potential underlying drivers and implications. We also discuss other features we included in our analysis, including multiple geographic indices, that were not found to have strong relationships with healthcare utilization patterns.

POTENTIAL POLICY OPTIONS

This paper concludes with a discussion of approaches policymakers can consider in light of these findings to address health disparities using Medicaid policy levers. We begin with considerations for making financial adjustments for social risk, either by adding new features to a risk-scoring algorithm or making equity adjustments outside of risk scoring. These approaches that better account for health disparities can mitigate disincentives to serve certain populations. This would more equitably distribute existing or additional resources to payers and risk-bearing providers than traditional risk adjustment, to better serve historically underserved populations. Additionally, we identify complementary policy approaches going beyond risk adjustment which states can consider to reinforce financial incentives, with a specific focus on leveraging state managed care contracts as well as broader policy options, including eligibility expansions, coverage of new benefits, workforce development, and engagement of people in communities most impacted by these disparities.

Risk adjustment

MECHANICS OF RISK ADJUSTMENT IN MEDICAID

In Medicaid managed care, managed care organizations (MCOs) contract with a state to provide healthcare services to Medicaid enrollees. Each MCO receives a per member per month (PMPM) capitation payment, which often varies by age, gender, geographic region, and reason for Medicaid eligibility. In many states, this capitation payment is further risk-adjusted to ensure payments reflect the health status of the MCO's members; MCOs enrolling a disproportionate share of members with higher risk, all else equal, receive higher payments, and vice versa for MCOs with lower-risk members. Most risk adjusters used for such a purpose are trained to predict individual claim costs. Some states may also incorporate other information into their risk adjustment process, such as social factors, gathered from outside the claim system.

BACKGROUND ON RISK SCORE ALGORITHMS

General

Risk adjustment comes in two forms: concurrent (also called retrospective) and prospective. The most well-known examples of these are the concurrent risk transfer program of the Affordable Care Act (ACA), and the prospective risk-adjusted capitation rate system used by the Centers for Medicare and Medicaid Services (CMS) in the Medicare Advantage program.

Concurrent risk adjustment programs make payments after the benefit year using a risk score that is based on the actual experience during the benefit year. These programs use a concurrent risk adjustment model, where the concurrent risk score represents an expected cost during the benefit year based on the person's clinical profile, as observed during the benefit year. Prospective risk adjustment systems make payments in advance or during the benefit year. These models use a prospective risk score, which predicts costs for an individual based on experience during a previous period.

Of course, risk score models can be used in applications other than risk adjustment transfers and payments. For example, concurrent risk adjustment models are often used in provider performance evaluation, experience studies, or retrospective value-based payments. Prospective risk scores are often used to predict future events, which can be useful in identifying members at high risk of adverse events, those in need of additional healthcare resources proactively, or in financial forecasts. Risk scores are often used to develop adjustments to capitation rates to account for program-wide shifts in acuity, such as occurred in many states due to rapid disenrollments following the end of the COVID-19 public health emergency (PHE).

Not all aspects of a person's health can be captured through medical claims; therefore, the use of other factors may lead to a more equitable distribution of funds. Our analysis will focus on the use of prospective models in identifying members likely to experience specific outcomes, and on how race, ethnicity, and social factors may impact the ability of a risk adjuster to make accurate predictions about the health status of various populations/cohorts.

CDPS overview

The Chronic Illness and Disability Payment System (CDPS) is a diagnostic classification system developed specifically to assist Medicaid programs with making capitation payments for the care of Medicaid beneficiaries.[2] It includes a suite of models designed for different eligibility groups, data inputs, and time periods. In addition to the diagnosis-based models, CDPS also includes a pharmacy-based model (MRx) and a combined diagnosis-and-pharmacy-based model (CDPS + Rx). All versions of the CDPS model include weights associated with various clinical conditions; each member's risk score is determined by adding up the weights for all conditions identified for the member. CDPS models are used by many state Medicaid agencies, often with customized weights calibrated to the state's specific population. This study uses the prospective CDPS + Rx model v7.0, both with the standard weights and a version customized to the study population.

2. See https://hwsph.ucsd.edu/research/programs-groups/cdps.html for more information on CDPS.

MARA overview

Milliman Advanced Risk Adjusters (MARA) is a risk adjustment system with a variety of models optimized for specific populations and applications.[3] MARA includes models developed using commercial health plan data and Medicare data but does not currently offer a model calibrated specifically to a Medicaid population. However, the MARA models that were calibrated using commercial health plan data have been successfully used in applications that involve Medicaid enrollees.

Another key difference between MARA and CDPS is in the scores that are output. CDPS outputs a total score, representing total member costs, while MARA outputs six score components in addition to the total: inpatient facility, outpatient facility, emergency department, physician, other medical, and retail pharmacy. In addition to the service category scores being useful for predicting costs within each category, the inpatient and emergency department scores have been found to be highly predictive of future inpatient admissions and emergency department visits, which are often of particular interest for organizations allocating care management resources.

This study uses MARA's prospective CxXPLN model. The MARA XPLN models are developed using regularized linear regression and behave similarly to other additive models, where each drug, condition, or demographic flag adds a value to the member's overall score. The CxXPLN model uses membership, medical, and pharmacy data, similar to the inputs used by CDPS + Rx, to make predictions.

MARA VS. CDPS PERFORMANCE COMPARISON

As part of the research for this paper, we compared the relative performance of MARA and CDPS on a Medicaid population. This study was based on data from two state Medicaid programs. Risk scores were developed, using diagnosis codes and pharmacy claims from July 2020 through June 2021, to predict costs during the July 2021 through June 2022 time period. We limited our analysis to full benefit populations, which excludes dual-eligible individuals receiving long-term services and supports (LTSS).

When comparing the performance of MARA and CDPS, it is important to keep in mind the differences in the models and their intended applications; MARA models have been optimized for high performance and are intended to be applicable for a wide variety of use cases, while CDPS models are designed for a specific managed care capitation use case, not necessarily to predict future individual costs with the highest degree of accuracy. Therefore, conclusions about the relative predictive accuracy of MARA compared to CDPS are not intended to indicate that MARA is more appropriate for use in all applications, particularly in applications related to setting Medicaid capitation rates. Furthermore, as noted below, MARA includes service category risk scores designed to predict the relative costs within specific types of service (inpatient, outpatient, professional, emergency department, pharmacy, and other medical), while CDPS models focus on total cost. It is also important to note that the customized CDPS model is being tested on the same data that was used to develop the weights, which may inflate some performance statistics relative to MARA and the baseline CDPS + Rx model, which have not been customized.

Because the focus of this research was on the use of specific types of services, we have focused our analysis of performance on the ability of the models to predict those services. In each of the comparisons shown below, we used MARA's service category risk scores alongside the same total risk score from the CDPS models for each category of service.

Coefficient of determination (R^2)[4]

We calculated R^2 for the baseline CDPS + Rx, custom CDPS, and MARA's CxXPLN service category risk scores, comparing risk scores for each aid category to the PMPM paid amount in the following year. Results are shown in Figures 1 through 4. As is shown in these figures, in all situations the MARA service category-specific score resulted in a higher R^2 (indicating better performance) than the CDPS models. The customized CDPS model had lower performance on most populations, likely because the parameters were chosen to minimize the root mean squared error (RMSE) rather than maximize R^2. We analyze results for the various eligibility categories separately because CDPS + Rx and the custom CDPS model use different models for different eligibility cohorts.

3. See https://us.milliman.com/en/risk/risk-adjustment for more information on MARA.

4. See https://en.wikipedia.org/wiki/Coefficient_of_determination for more information on R2.

The higher predictive performance of MARA's service category scores indicates they are likely better suited to control for the impact of differences in clinical factors on the services of interest; however, additional study is needed to confirm this hypothesis.

FIGURE 1: R^2 FOR PROSPECTIVE INPATIENT (IP) PAID PMPM

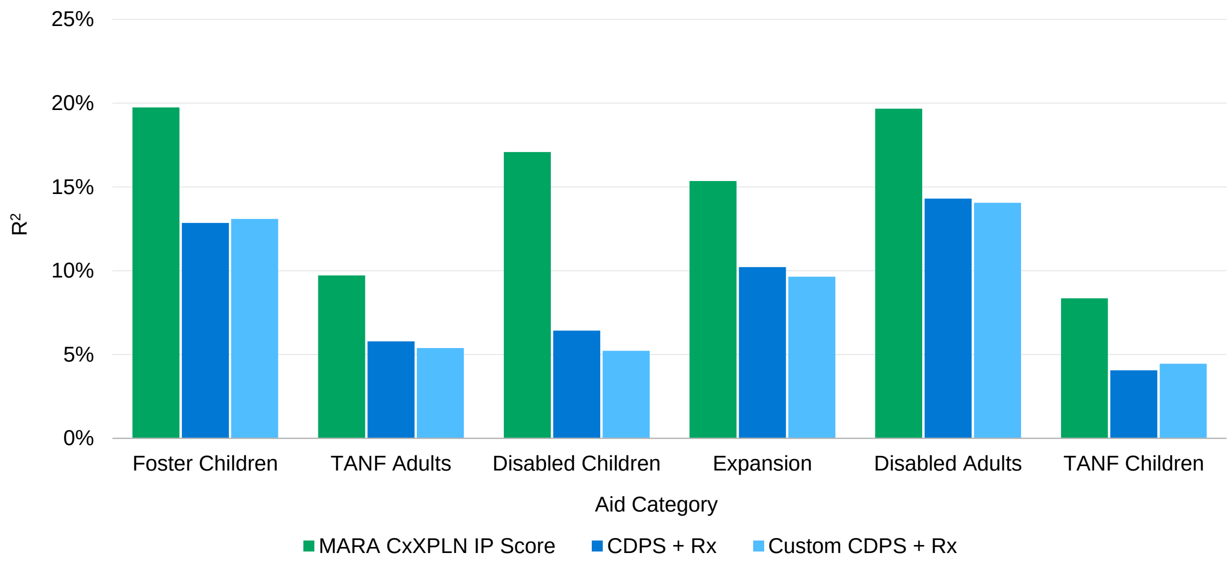


FIGURE 2: R^2 FOR PROSPECTIVE EMERGENCY DEPARTMENT (ED) PAID PMPM

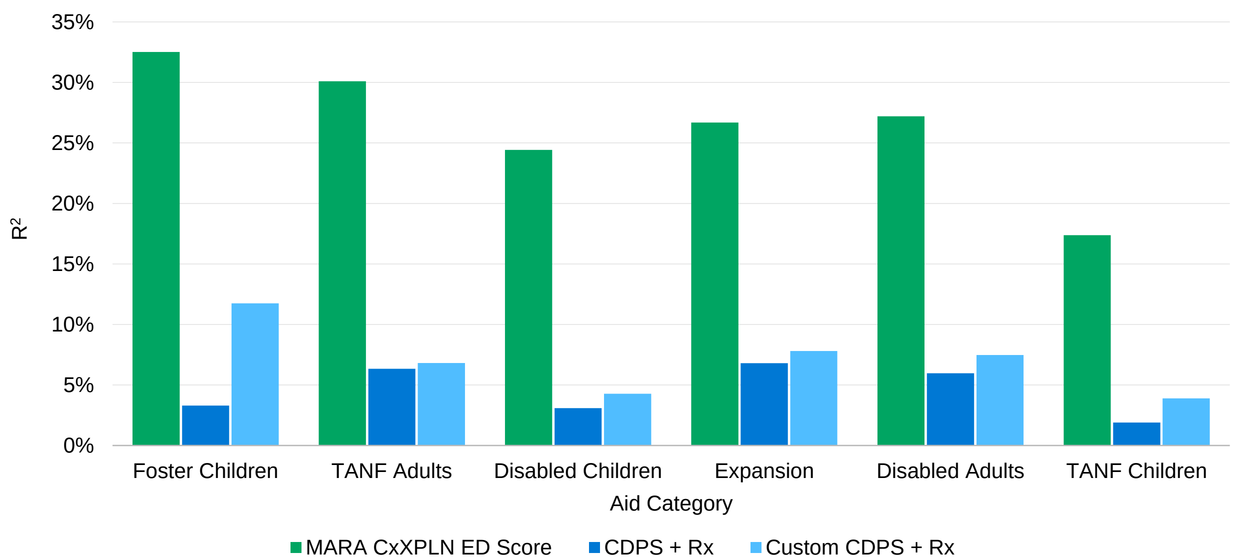


FIGURE 3: R^2 FOR PROSPECTIVE PHYSICIAN (PHY) PAID PMPM

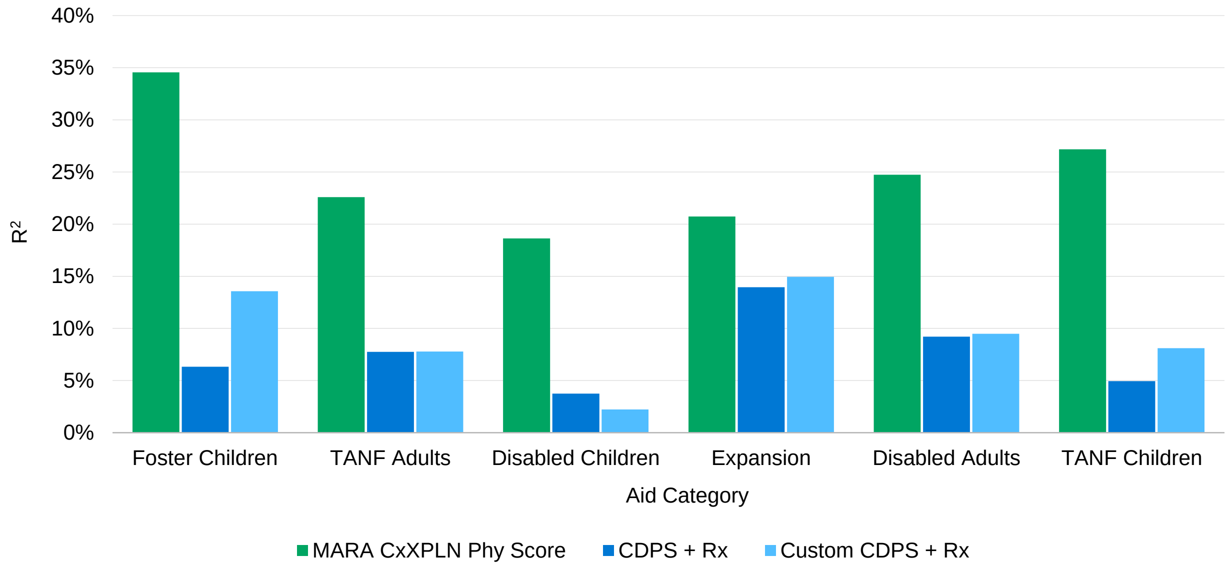
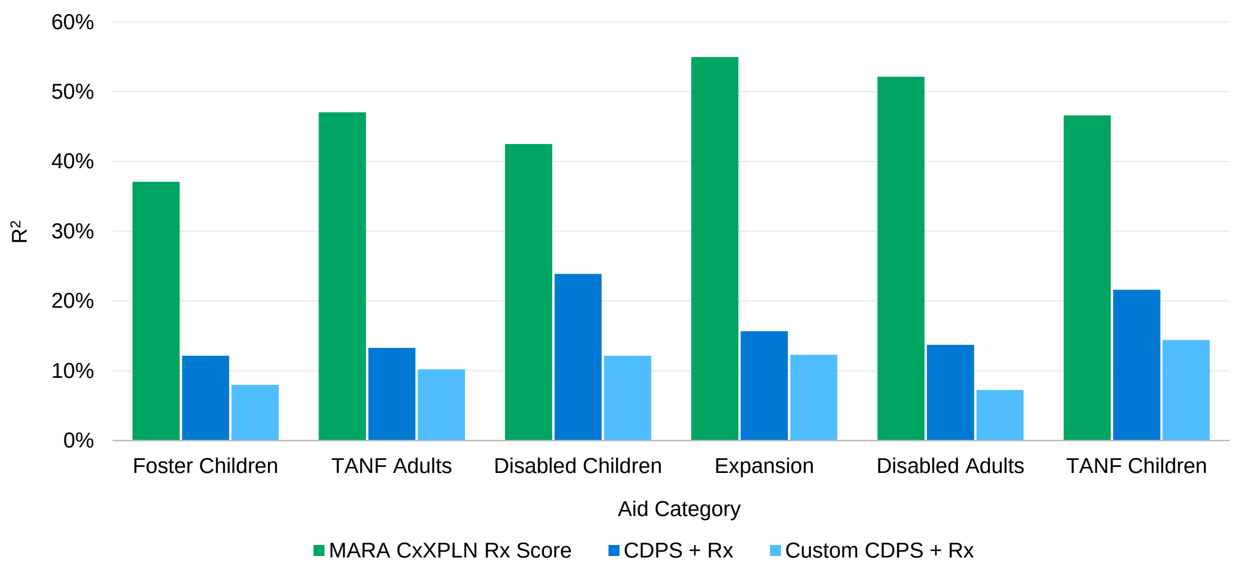


FIGURE 4: R^2 FOR PROSPECTIVE PHARMACY (RX) PAID PMPM



Event probabilities

Another way to measure the predictive power of a model, particularly suited to inpatient and emergency department utilization, is to examine the likelihood of having an inpatient admit or emergency department visit at different risk score levels. We summarize this information for MARA and the custom CDPS + Rx model in Figures 5 through and 6 (note, performance for the custom CDPS + Rx model was found to be similar to the non-customized CDPS + Rx model and therefore only the custom model is shown below). The triangles and squares shown on the figures represent the likelihood of an inpatient admit or emergency department visit for the custom CDPS + Rx model and for the corresponding MARA service category score, respectively. The bars show the average inpatient or emergency department cost for the utilizers within that risk score strata. The graphs show that stratifying the population using the MARA service category risk score provides a better ability to focus on the members who are most likely to utilize these services, and the higher average cost indicates an improved ability to identify frequent utilizers or more severe events. It is important to keep in mind while reviewing these metrics that neither model is designed to predict utilization, though risk scores are often used to predict or normalize utilization of specific services.

FIGURE 5: PROSPECTIVE INPATIENT ADMIT PROBABILITIES AND COSTS PER UTILIZER BY RISK SCORE

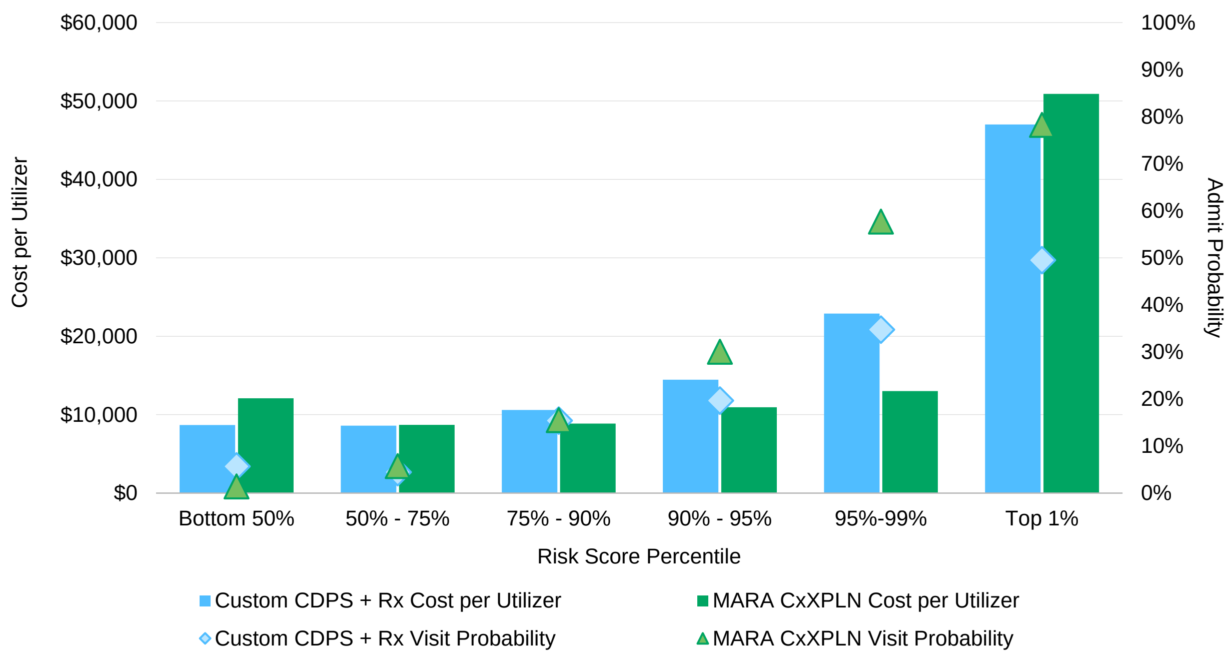
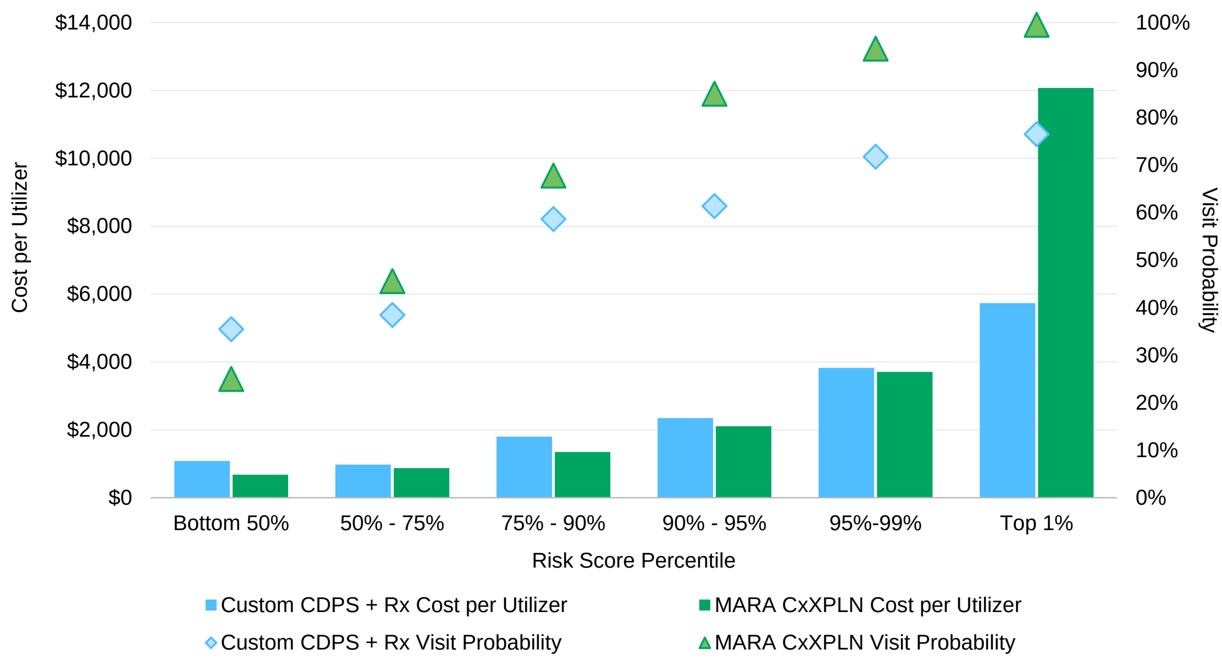


FIGURE 6: PROSPECTIVE EMERGENCY DEPARTMENT VISIT PROBABILITIES AND COSTS PER UTILIZER BY RISK SCORE



Given the improvement in performance associated with MARA service category risk scores, we chose these risk scores as the means to control for clinical factors to better isolate the impact of social and demographic factors on utilization.

COMMENTARY ON CUSTOMIZING RISK ADJUSTMENT ALGORITHMS

Often, users will create custom risk adjustment models or other predictive models for specific purposes and populations. Doing so can be extremely valuable, and such an exercise may improve the ability of the risk adjustment models to account for the impact of clinical factors on utilization patterns in this research. For example, a model specifically designed to predict the number of inpatient admissions may have performed better than using the CDPS or MARA risk scores. However, customization comes with its own set of challenges, and organizations do not always have the resources to build or maintain different models for different purposes. Building custom models requires a large amount of data to ensure results are credible and generalize to populations other than the training data. Models like CDPS and MARA are usually built using experience from several million lives, which exceeds the size of data available for most individual state Medicaid programs, particularly because prospective models generally require eligibility in both the assessment period and a future projection period. Medicaid programs often have significant member turnover, limiting the size of the longitudinal population that can be used to train a prospective risk score model. Another challenge is the time, effort, expertise, and computing resources necessary to train models. Getting data into the necessary format and creating and testing models requires specialized expertise and significant resources.

Due to these limitations and challenges, many organizations use one model for multiple purposes, similar to the use of MARA in our research.

Analysis of healthcare utilization patterns

STUDY DESIGN AND VARIABLES ANALYZED

In the next phase of our research, we sought to understand how certain social, demographic, and economic factors were related to utilization of healthcare services. This subsection discusses the population and variables included, and the following subsection focuses on discussion of the most pertinent findings. Additionally, Appendix A has further technical details related to the model and feature selection, theoretical framework, and overall analytical approach.

Population and time periods included

Consistent with the risk score study described in the previous section, this research was based on Medicaid managed care encounter data from two states. We limited our analysis to full benefit populations, which excludes dual-eligible individuals receiving LTSS. We identified features in the period July 2020 through June 2021. We measured response variables in the period July 2021 through June 2022, although there were some modifications for maternity visits, which are discussed in the description of those variables. We further limited the analysis to members with enrollment in all 12 months in the feature period and at least one month of enrollment in the response period.

Response variables

In the comparison of MARA versus CDPS, discussed in the prior section, we focused on predictions of PMPM expenditures, which aligns with how these risk score algorithms are typically used in Medicaid programs. However, in this second phase of our research, we were interested in gaining a better understanding of care patterns, which may be obscured by using PMPM expenditures as the response variable. We ran separate models to predict the following response variables:

- **Inpatient medical/surgical hospital admissions.** We excluded maternity, psychiatric, and rehabilitation admissions to focus on acute medical events.
- **Emergency department (ED) visits:** We relied only on hospital claims (rather than physician claims) to identify emergency department visits. Observation stays were counted as emergency department visits, but emergency department visits that resulted in inpatient admissions were excluded from this variable and were instead counted as inpatient admissions.
- **Primary care physician (PCP) office visits:** We recognize that there is a wide range of possible definitions of “primary care,” but we included both in-person and telehealth visits with a physician or mid-level clinician, such as a nurse practitioner or physician assistant. We restricted physicians to those with certain provider taxonomy codes, such as those related to internal medicine, family medicine, or pediatrics. We also counted all medical visits to a federally qualified health center (FQHC) or rural health clinic (RHC). The full list of provider taxonomy codes, Current Procedural Terminology (CPT) codes, and Healthcare Common Procedure Coding System (HCPCS) codes used to identify PCP office visits is included in Appendix B.
- **Retail pharmacy scripts:** We counted each claim as a single script regardless of the cost, unit count, or days’ supply. We excluded drugs administered in a hospital or office setting. In addition to total retail pharmacy scripts, we also separately included response variables for two particular therapeutic classes that had been shown in preliminary analyses to vary widely by race and ethnicity.
  - **Antidepressants:** Identified using the Wolters-Kluwer Medi-span databases.
  - **Antianxiety agents:** Identified using the Wolters-Kluwer Medi-span databases.

- **Maternal care visits:** We focused separately on prenatal and postpartum visits. We limited our analysis to completed pregnancies, so it was important to first identify births in the data. We identified using ICD-10 and CPT codes that indicated birth, as well as CPT codes related to delivery, using claims from all settings of care to ensure comprehensive identification. We then used the second six months of the study period to identify births, ensuring we had sufficient data for observing prenatal visits.
  - **Prenatal visits:** After identifying births within the study period, we utilized a nine-month look-back period to capture prenatal visits. Each medical visit with a unique date, identified with a prenatal-related ICD-10 diagnosis code, was counted as a unique prenatal visit.
  - **Postpartum visits:** We identified postpartum visits using HCPCS, CPT, and ICD-10 diagnosis codes. We employed the same comprehensive approach, using claims from all settings of care. Typically, the postpartum period is considered to extend up to six weeks after delivery. However, for comprehensive healthcare analysis, we extended the postpartum tracking period up to 12 months to ensure all relevant postpartum care is captured. Due to the availability of data, for births identified at the end of the study period, this provided approximately two months of postpartum tracking. This timeframe still ensures that at least the standard six weeks of postpartum care is covered. We counted each medical visit (identified with a postpartum-related ICD-10 diagnosis code, HCPCS code, or CPT code) with a unique date as a unique postpartum visit.

Feature variables

Our analysis included a wide range of characteristics that we hypothesized may be related to healthcare utilization patterns.

- **Risk scores:** We used MARA service category scores. These scores represent predictions for costs related to certain service categories; for each of the response variables listed above, we selected the most applicable MARA service category prediction.
- **Z-codes in healthcare data:** ICD-10 diagnosis codes starting with “Z” are used to denote a range of social determinants of health (SDOH) that affect an individual's overall well-being and healthcare needs. These codes capture information beyond medical diagnoses, encompassing various social, economic, and environmental factors. Z-codes include categories such as housing instability, employment issues, social environment, and other significant life circumstances that impact health outcomes. Integrating Z-codes into healthcare data analysis provides a more comprehensive understanding of the factors influencing patient health and can inform targeted interventions and policy decisions.
- **Homelessness indicator:** To identify people experiencing homelessness within the study population, we employed a multistep approach. We utilized the Z59.0 ICD-10 diagnosis code, specifically designated for homelessness, as our primary indicator. Additionally, we incorporated geocoding of shelter addresses and the residential addresses from eligibility files. By mapping the coordinates of known shelter locations and matching them with the residential addresses of individuals, we identified likely to be residing in shelters. We further refined this geographic analysis using natural language processing to match names of shelters with the data, ensuring that cases where the shelter name was used (instead of an address) were also captured. We also included self-identified cases where individuals explicitly reported themselves as homeless.

  For one of the states in our study, we leveraged historical address data available to flag individuals who had ever been homeless. This longitudinal approach allowed us to capture the episodic nature of homelessness. For the second state, with only one year of data available, we identified homelessness based on the information within that timeframe. Ultimately, by combining the Z59.0 code, address information, and self-reported data, we created a robust methodology to identify people experiencing homelessness within the study population. In aggregate, we identified approximately 0.6% of the study population as homeless.

- **Other Z-codes:** Other Z-codes denote a range of social factors affecting an individual. We grouped them into 10 categories, each included as a feature in the study. The ICD-10 diagnosis code indicating homelessness (Z59.0) was excluded from these categories because it was broken out as a separate feature described above. A limitation inherent with Z-codes is that they must be coded on a medical claim, meaning they only record information about members using the healthcare system. Therefore, the members coded with these conditions likely only reflect a subset of the members affected by these social factors. However, we determined that the prevalence of these codes was sufficiently large and stable to include in the study. In each year of our study, we observed roughly 2% of members coded with at least one of these Z-codes (other than homelessness). In the Potential Policy Options for Medicaid Agencies section of this report below, we discuss considerations for states as they think about implementing these codes into the risk adjustment process.
- **Race/ethnicity:** We relied on self-reported race and ethnicity values from each state, supplemented by an imputed race/ethnicity prediction for members who did not report a race. Approximately 65% of members self-reported a race/ethnicity value. For the remaining 35% of members, we applied a Bayesian Improved Surname Geocoding (BISG) algorithm[5] to estimate the race based on first name, last name, and member address. The BISG algorithm outputs probabilities for each potential race/ethnicity group. For purposes of our analysis, we allowed members to have partial membership in multiple race/ethnicity groups based on these probabilities. For example, a member with a 75% probability of being white and 25% probability of being Black received a value of 0.75 for the “white” field and 0.25 for the “Black” field, and therefore their results would contribute to both the Black and white cohort results. By contrast, a member with a self-reported race of Black received a value of 1.00 in the “Black” field and a 0 in all other race/ethnicity fields. The available values for self-reported race/ethnicity varied slightly by state and the BISG algorithm, so we rolled the values up to the following:
  - Asian
  - Black
  - Hispanic
  - White
  - Multiracial or other
  - Unknown
- **Rate cell:** In both of the states in our analysis, members are grouped into rate cells for purposes of capitation rate setting. The rate cells are generally based on a combination of the member’s reason for Medicaid eligibility, such as Temporary Aid for Needy Families (TANF), Supplemental Security Income (SSI), Medicaid expansion, or foster child, along with age and gender. Because the rate cell groupings varied between the two states, we used detailed eligibility, age, and gender data for each member to create standardized rate cells for this analysis. Some states use a separate “kick payment” rate cell that covers maternity services and is triggered by a delivery claim. For purposes of this analysis, we included maternity claims within the member’s original rate cell and did not create a separate kick payment rate cell. The full list of rate cells included in this analysis is shown in Appendix C.
- **Mental health and substance use disorder (SUD) count:** We used ICD-10 diagnosis codes to count the number of unique mental health and SUD conditions identified for each member. Each unique ICD-10 diagnosis code was treated as a separate condition for purposes of this variable.

5. For considerations related to the use of race and ethnicity imputation, see the paper “Statistical Methods for Imputing Race and Ethnicity” from the Society of Actuaries, published in April 2024, available at https://www.soa.org/498d64/globalassets/assets/files/resources/research-report/2024/stat-methods-imputing-race-ethnicity.pdf.

- **Social Deprivation Index (SDI) components:** Based on each member's address, we mapped based on the seven components of the 2019 SDI, which is maintained by the Robert Graham Center.[6] The SDI is used to quantify socioeconomic variation in health outcomes. These components are derived from the American Community Survey and vary by census tract. The seven components are:
  - Percentage of population with income under 100% of the federal poverty level (FPL)
  - Percentage of population aged 25 years of age or older with less than 12 years of education
  - Percentage of population aged 16-64 years who are non-employed
  - Percentage of households living in renter-occupied housing units
  - Percentage of households living in crowded housing units
  - Percentage of families that have a single parent with dependents under 18 years of age
  - Percentage of households living with no vehicle
- **Rural vs. urban indicator:** Based on each member's address, we used the 2023 Rural-Urban Continuum Codes (RUCCs) to classify members as being either rural or urban. The RUCCs are maintained by the U.S. Department of Agriculture (USDA), Economic Research Service (ERS).[7] The RUCCs assign each county a code from 1 to 9, with 9 having the smallest population. For our analysis, we treated codes 4 through 9 as rural, which is consistent with other healthcare research.[8]
- **Healthcare provider shortage area (HPSA) score:** Based on each member's address, we mapped on this value from the Health Resources and Services Administration (HRSA).[9] The specific HPSA score we used indicates the level of primary care shortage experienced in a given geographic area.
- **Nutrition access indicator:** Based on each member's address, we applied the "LA 1 and 20" index from the Food Access Research Atlas, which constructs a binary number for census tracts to identify areas with limited access to affordable and nutritious food. Low access indicates that a significant share of residents in the tract are more than one mile (urban areas) or 20 miles (rural areas) from the nearest supermarket.

The MARA risk scores account for a wide range of clinical factors, and therefore we largely focused on nonclinical factors for our other features. In some cases, we included certain variables (such as a count of mental health diagnoses) that overlap with information included in the MARA risk score but which we believe may have an outsized effect on the population and response variables in this analysis. It is possible that, if we had fully recalibrated the MARA risk score on a Medicaid population or to predict the specific categories of interest, the risk score would have inherently reflected some of these features.

FINDINGS

The goal of our analysis was to understand the relationships between feature and response variables. Given that we had nearly 60 total feature variables and eight response variables, the models produced a high volume of outputs. We reviewed the outputs in detail to identify the most notable findings, which we have summarized below. This exercise required a certain level of judgment. In general, we focused on instances where a feature resulted in substantially higher or lower predictions for a certain response variable, or where we identified a pattern of interest across multiple response variables (such as lower PCP visits paired with higher emergency department visits). It is possible that additional conclusions could be drawn from the output of the models.

It is important to note that our findings are primarily based on the outputs of the models, rather than summary statistics of the data. The intent is to isolate the effect of each feature independently, assuming all other features are held constant.

6. Robert Graham Center. Social Deprivation Index (SDI). Retrieved November 26, 2024, from https://www.graham-center.org/maps-data-tools/social-deprivation-index.html.
7. USDA ERS. Rural-Urban Continuum Codes. Retrieved November 26, 2024, from https://www.ers.usda.gov/data-products/rural-urban-continuum-codes.aspx.
8. Leblanc, G. et al. (February 14, 2022). Rural-Urban Differences in Breast Cancer Stage at Diagnosis. Womens Health Rep (New Rochelle). Retrieved November 26, 2024, from https://pubmed.ncbi.nlm.nih.gov/35262058/.
9. HRSA. Find Shortage Areas. Retrieved November 26, 2024, from https://data.hrsa.gov/tools/shortage-area.

Race/ethnicity

We observed material variation by race/ethnicity for each of the response variables in the analysis. At a high level, we drew several key conclusions, which we will provide support for later in this section.

1. Black individuals had higher utilization of emergency department visits and inpatient admissions, but lower utilization of primary care visits and retail pharmacy scripts. This finding suggests that a lack of routine care and ambulatory services may be resulting in greater need for acute care for this cohort.
2. White individuals generally had higher utilization of all services, most notably retail pharmacy. Utilization of antidepressants and antianxiety agents were often two to three times higher than for other race/ethnicity groups. This suggests there may be differences in provider prescribing patterns, patient treatment preferences, or cultural influences.
3. Hispanic individuals had the highest utilization of primary care visits but the lowest utilization of retail pharmacy scripts. This is a somewhat counterintuitive finding, given that primary care physicians are responsible for prescribing many medications. It is possible that Hispanic individuals are prescribed medications less frequently, fill prescriptions less frequently, or a combination of both.
4. Asian individuals have the lowest utilization of emergency department visits and inpatient admissions, but roughly average utilization of primary care visits and retail pharmacy scripts. These relationships suggest that appropriate use of ambulatory care may be mitigating the need for acute services for Asian individuals.

The graphs in Figures 7 to 10 illustrate how the predictions for a given response variable are affected by the member's race/ethnicity. In each of these graphs, all other features are held constant at the average level across our sample so that we can isolate the effect of race/ethnicity on the predictions. For these visualizations, we have grouped response variables into pairs that exhibited similar relationships. The percentages next to each race/ethnicity label on the horizontal axis indicate the percentage of enrollees in the sample data that were in that group. The values are different in the chart related to maternity services, because that sample was limited to deliveries only.

Figure 7 shows emergency department visits (gray bars) and inpatient admissions (blue bars), both of which are high-cost services typically driven by acute health events. The relationships with race/ethnicity were similar for these services: utilization was higher for Black and white individuals and lower for Asian and Hispanic individuals.

FIGURE 7: PERCENTAGE CHANGE IN PREDICTIONS DUE TO RACE/ETHNICITY, EMERGENCY DEPARTMENT VISITS AND INPATIENT ADMISSIONS

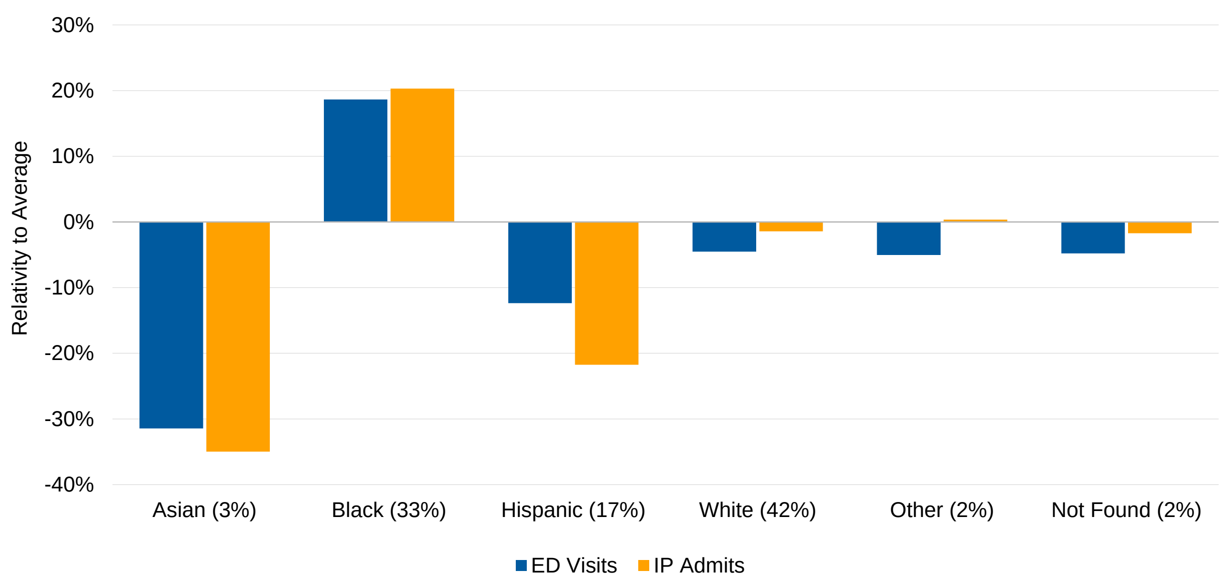
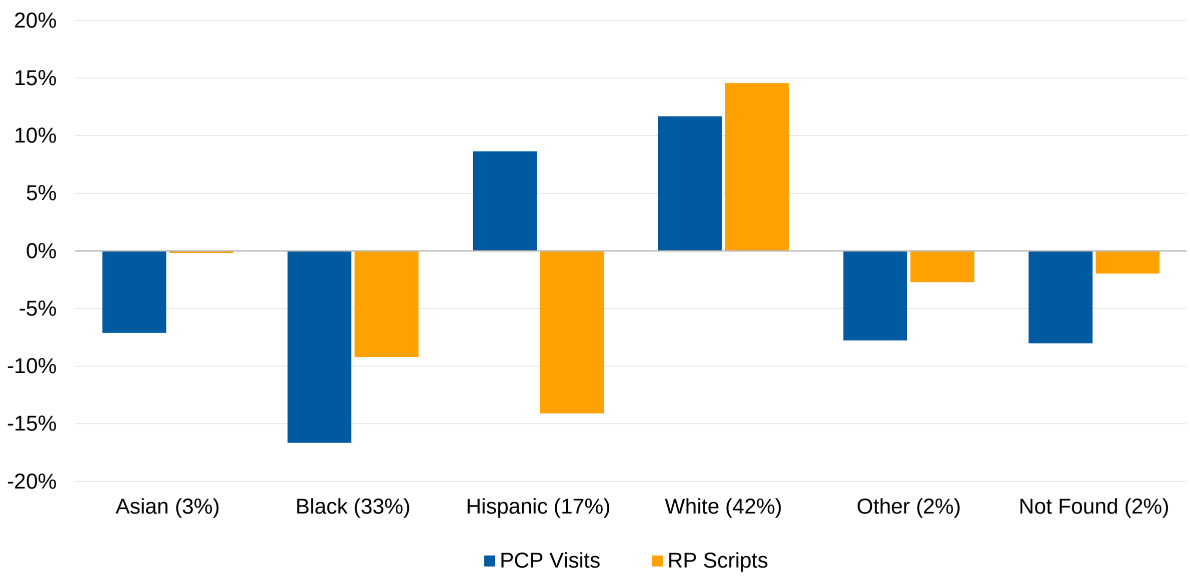


Figure 8 illustrates the variation by race/ethnicity for PCP visits and retail pharmacy scripts. With the exception of the Hispanic population, we observed similar relationships for the other race/ethnicity groups in Figures 7 and 8. Utilization was low for the Black and Asian populations and higher for the white population. For the Hispanic population, utilization was second-highest among all race/ethnicity groups for PCP visits, but lowest for retail pharmacy scripts.

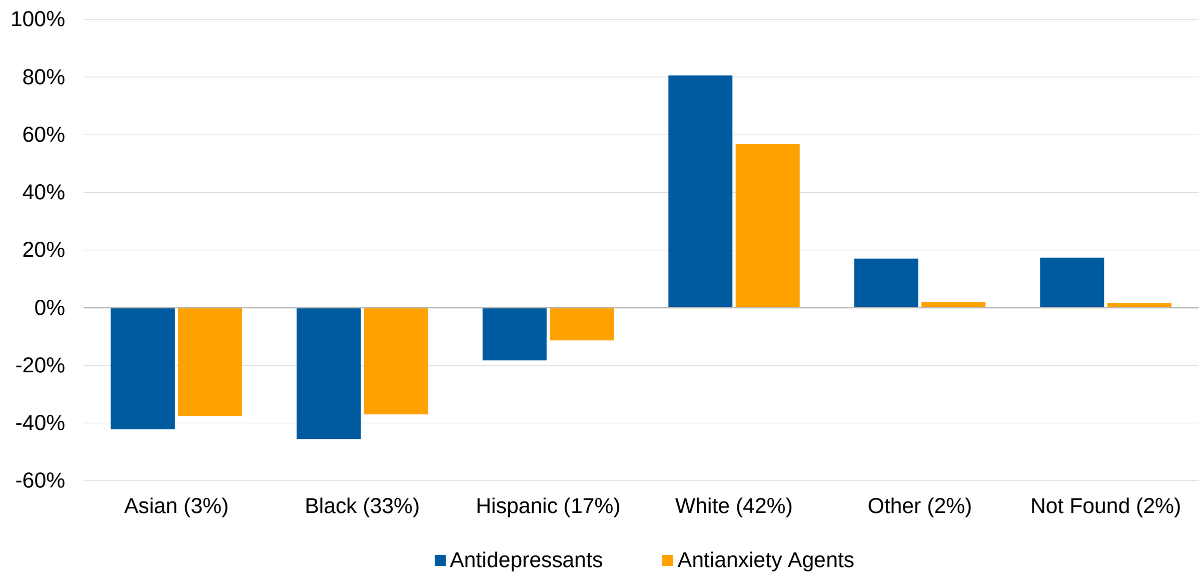
FIGURE 8: PERCENTAGE CHANGE IN PREDICTIONS DUE TO RACE/ETHNICITY, PCP VISITS AND RETAIL PHARMACY SCRIPTS



Digging deeper into the retail pharmacy category, Figure 9 shows the variation in predictions for antidepressants and antianxiety agents. Here we see striking differences between the white population and all other race/ethnicity groups. For antidepressants, utilization for white individuals was more than 80% above the average, and more than three times as high as for Black and Asian individuals. This implied disparate use of antidepressants is consistent with a 2020 data brief from the Centers for Disease Control and Prevention (CDC), which found that the prevalence of antidepressant use was more than twice as high among non-Hispanic white adults as other race/ethnicity groups.[10]

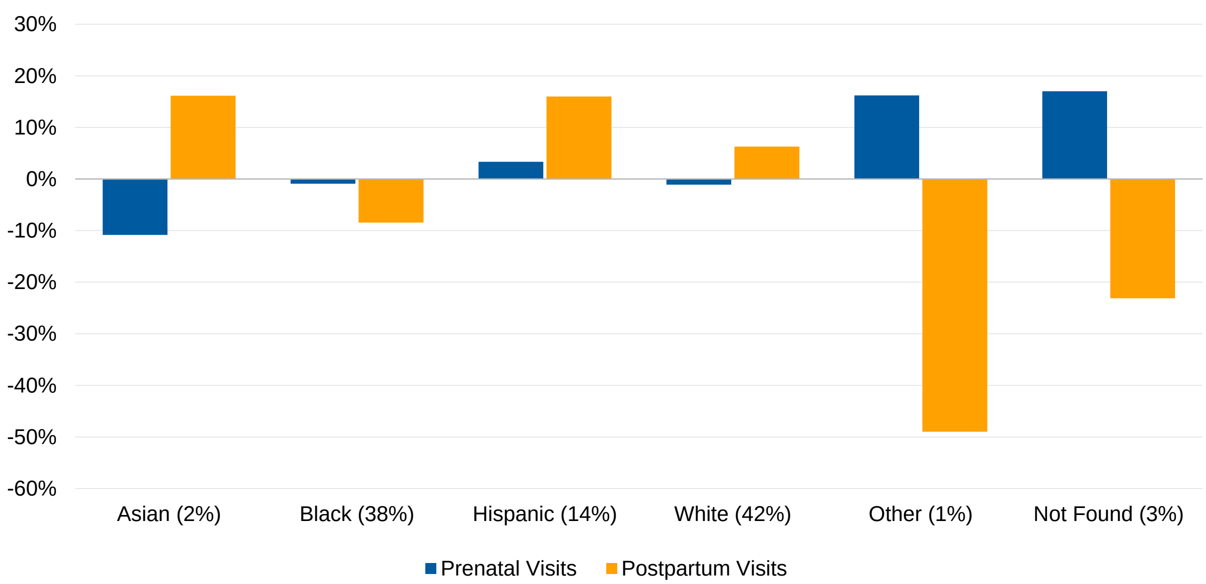
10. Brody, D.J. & Gu, Q. (September 2020). Antidepressant Use Among Adults: United States, 2015-2018. CDC. Retrieved November 26, 2024, from https://www.cdc.gov/nchs/products/databriefs/db377.htm.

FIGURE 9: PERCENTAGE CHANGE IN PREDICTIONS DUE TO RACE/ETHNICITY, ANTIDEPRESSANT AND ANTIANXIETY SCRIPTS



Lastly, Figure 10 illustrates prenatal and postpartum visits. In general, we observed relatively little variation among the race/ethnicity groups in terms of prenatal visits. However, for postpartum visits, individuals in the “Other” or “Not Found” cohorts had markedly lower utilization than other groups. Unfortunately, it is more difficult to draw conclusions for these two cohorts, as they are comprised of a variety of smaller subsets, including American Indian, Pacific Islander, and multiracial individuals, as well as individuals who did not self-report a race/ethnicity or address (and therefore a race/ethnicity could not be imputed).

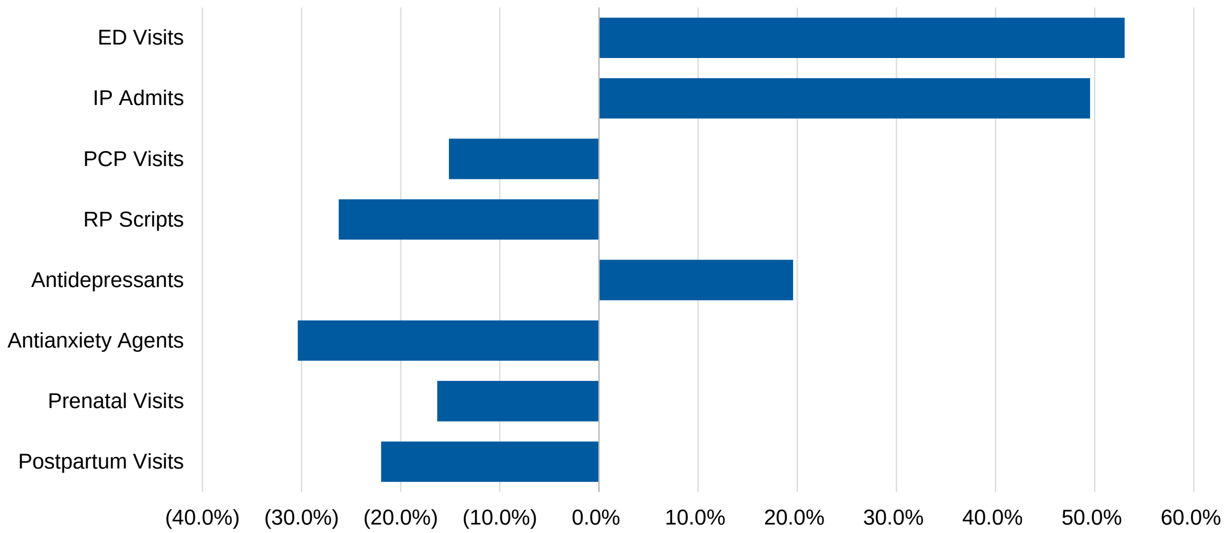
FIGURE 10: PERCENTAGE CHANGE IN PREDICTIONS DUE TO RACE/ETHNICITY, PRENATAL AND POSTPARTUM VISITS



Homelessness

The next feature that appeared to be a significant driver of differences in utilization patterns was homelessness. Figure 11 illustrates the percentage change in the predictions for each of our response variables when an individual was identified as experiencing homelessness. People experiencing homelessness were found to have roughly 50% more inpatient admissions and emergency department visits compared to similar individuals who were not homeless. However, utilization of most other services was predicted to be lower for people experiencing homelessness. This strongly suggests that people experiencing homelessness are less likely to access ambulatory care, which might have been able to mitigate their increased use of acute care in the hospital.

FIGURE 11: PERCENTAGE CHANGE IN PREDICTIONS DUE TO HOMELESSNESS

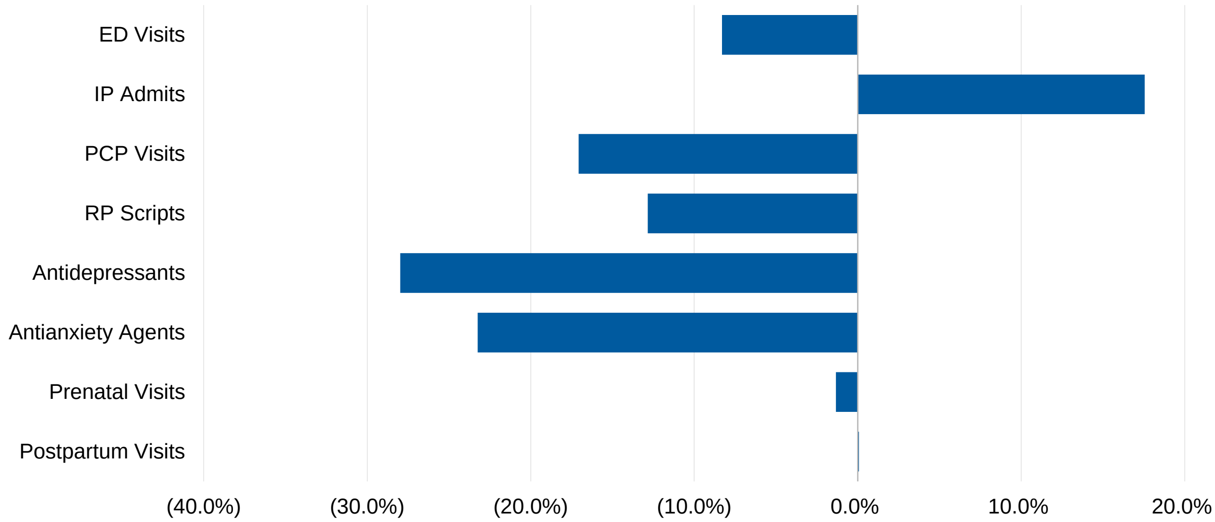


Because of the difficulties in defining homelessness and identifying individuals meeting this criterion, it is possible that the actual disparities related to homelessness are even greater than we have shown in our analysis. In particular, we identified the presence of homelessness in the period prior to the measurement year, consistent with other feature variables. However, individuals who first experienced homelessness during the measurement year are not flagged as homeless in the analysis. It is likely that these individuals also had material disparities in utilization patterns, but they are grouped with the housed population in this analysis.

Rural vs. urban

We also observed notable differences in utilization patterns when individuals lived in rural areas, as shown in Figure 12. Individuals in rural areas had lower utilization (compared to urban areas) of all services in our analysis with the exception of inpatient admissions, which were nearly 20% higher. This finding could be due to challenges accessing ambulatory care due to long driving distances and fewer providers per capita. The higher inpatient admissions may result from: 1) Care that has been postponed until the severity required admission to a hospital, or 2) The lack of availability of a lower level of care that could address certain situations.

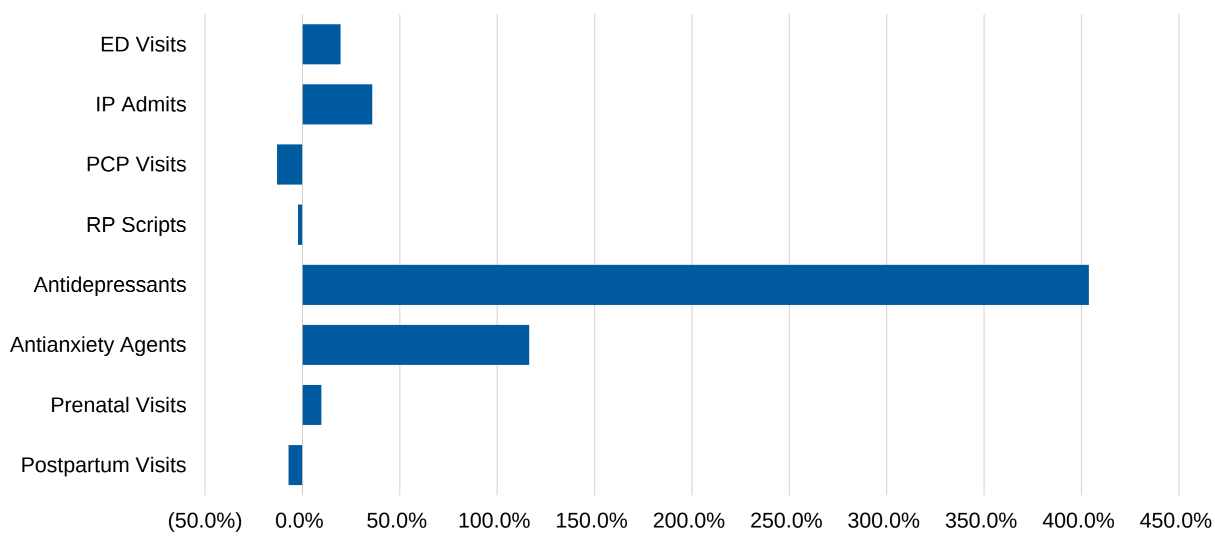
FIGURE 12: PERCENTAGE CHANGE IN PREDICTIONS DUE TO LIVING IN A RURAL AREA



Problems related to upbringing

Aside from homelessness, the Z-codes that had the greatest disparities in utilization were codes grouped under "problems related to upbringing." This includes ICD-10 codes Z62.2 through Z62.892, with diagnoses such as being in custody of a non-parental relative or a history of parental abuse. Figure 13 shows the percentage change in the predictions for each of our response variables when an individual had been flagged with one of these conditions. Most obvious is the extremely high utilization of antidepressants and antianxiety agents, but emergency department visits and inpatient admissions were both materially higher for these individuals than for similar individuals who were not identified with these conditions. Interestingly, total retail pharmacy scripts were slightly lower for these individuals, which would imply relatively low utilization of medications not related to mental health. It is worth noting that 75% of the roughly 16,000 individuals in our data flagged for these conditions were children. Both the child and adult cohorts had rates of antidepressant utilization that were more than three times the population-wide average of approximately 0.62 per enrollee per year.

FIGURE 13: PERCENTAGE CHANGE IN PREDICTIONS DUE TO PROBLEMS RELATED TO UPBRINGING



SDI and other geographic indices

As described in the Study Design and Variables Analyzed section above, we included a wide range of features in our model. In general, features not covered earlier in the Findings subsection had relatively weak relationships with the response variables we analyzed. We will not cover all these features in detail. However, we did find it notable that, other than the rural/urban indicator, the other geographic features included in our analysis had such modest effects. For instance, the predicted PCP visits, emergency department visits, and inpatient (IP) admissions were less than 3% different for individuals residing in an area with a PCP shortage, as defined by HPSA. As another example, when comparing individuals in areas with the lowest income quartile to those in areas with the highest income quartile, the predictions for each response variable differed by no more than 9%. When reviewing all modeled coefficients for these geographic features, 85% of the coefficients were within +/- 0.10.

While these findings are not sufficient to conclude that there was no variation in utilization of health care services by geography in our data, they do suggest that these differences were largely explained by other individual-level factors, such as age, gender, disability, health status, race/ethnicity, and homelessness. Additionally, the weak findings related to the HPSA also suggest that challenges related to “access” may be more complex than can be explained by geographic-level data. Contrasting this with the wider variation by race/ethnicity, it is possible that, even if providers are available in an area, there may be racial, ethnic, or cultural differences between the providers and certain groups in the community.

Potential policy options for Medicaid agencies

RISK ADJUSTMENT

In recent years, there has been growing interest in exploring the addition of social risk factors and demographic indicators in risk adjustment methodologies as a strategy to improve equity in the distribution of healthcare dollars and to improve incentives to reach historically underserved populations.[11] However, the relationship between these variables and healthcare spending is often not straightforward. As demonstrated in this analysis, historically underserved racial/ethnic groups may experience higher utilization of certain services and lower utilization of others, which may result in similar or even lower healthcare spending overall. For example, observed disparities in primary care utilization should not be construed to suggest Black and Asian individuals need fewer primary care services than their white and Hispanic counterparts. Simply adding race/ethnicity to a risk-scoring algorithm as a predictor of healthcare spending can lead to an underinvestment in certain communities whose relatively low utilization may not necessarily be a reflection of better health or lower healthcare needs, but rather impediments to access or less complete coding of underlying diagnoses.

The complexity and ethics surrounding social risk adjustment demands a nuanced and thoughtful approach to considering how to incorporate these characteristics into healthcare financing to make sure that incentives are structured in a way that promotes investment in and support of communities that experience disparate health outcomes. Policymakers need to invest the time and effort up-front to conduct nuanced analyses to understand relationships between risk factors and health outcomes. For example, our analysis found that the SDI had relatively weak relationships with healthcare utilization. However, when we looked at the impact of specific member-level characteristics, such as member race/ethnicity or homelessness, we found stronger relationships. Although potentially influential variables such as housing, income, and education are incorporated into the SDI, they are aggregated by geography, which limits their predictive power. Having a more complete understanding of social risk factors and disparate patterns of utilization allows policymakers to explore whether it is appropriate and equitable to include these factors within a risk score or to look for other ways to adjust payments in order to drive resources to eliminate those disparities.

Adjustments to risk scores

For factors that have a clear and clinically reasonable impact on health spending, like homelessness or problems related to upbringing (both of which were shown in this analysis to be associated with higher utilization of many acute services), inclusion of the factor in risk scoring may be an effective tool to enhance managed care financing by mitigating potential incentives to avoid covering these populations and providing resources to address the conditions that are driving their utilization. There is precedent for incorporating indicators of homelessness and housing instability into risk adjustment. For example, Massachusetts and Minnesota have added housing instability indicators in risk adjustment in recent years using Z-codes, addresses associated with homeless shelters, or more than three addresses on file in one year.[12] States interested in adding social indicators in risk adjustment for MCO payments may wish to consider also developing managed care contract requirements aimed at improving care and health outcomes for the specific populations of interest. These strategies are discussed further as we discuss policy approaches beyond risk adjustment below.

Adjustments outside of risk scores

For social risk factors or demographic characteristics where disparities are known to exist but the relationship with healthcare spending may be less straightforward, we advise caution when considering incorporation of these factors as a predictor in a risk score algorithm to avoid unintentionally driving investment away from populations whose

11. Phillips Jr, R.L. et al. (March 27, 2023). Accounting For Social Risks In Medicare And Medicaid Payments. Health Affairs. Retrieved November 26, 2024, from https://www.healthaffairs.org/content/forefront/accounting-social-risks-medicare-and-medicaid-payments.

12. University of Minnesota State Health Access Data Assistance Center (August 2020). Risk Adjustment Based on Social Factors: State Approaches to Filling Data Gaps. Retrieved November 26, 2024, from https://www.shvs.org/wp-content/uploads/2020/08/FINAL_SHVS-Risk-Adjustment-Brief.pdf.

relatively lower healthcare spending may be a result of inequities in access. As an example from our research, we found that enrollees in rural areas had generally lower utilization of many services, which may be related to challenges in accessing care due to driving distances. Incorporating a rural indicator into a risk score algorithm may have resulted in lower risk scores for members residing in those areas, and therefore less capitation rate funding would be directed to MCOs serving those enrollees. This reduced funding would make it more difficult for MCOs to allocate resources toward connecting these enrollees with the care they need. In all cases, one should have a comprehensive understanding of the impact any included social factors will have on the risk score to ensure that the downstream effects on covered individuals and MCOs are consistent with the intended purpose of incorporating the information.

That said, there are also ways to apply adjustments outside of a predictive risk score to enhance the financing of MCOs or risk-bearing provider entities that serve populations with higher social needs. With this kind of approach, the goal of risk adjustment shifts from serving as purely a predictor of future spending to also accounting for expected resources needed to engage historically under-resourced communities. A notable example is the Health Equity Benchmark Adjustment (HEBA) used in CMS's Accountable Care Organization (ACO) Realizing Equity, Access, and Community Health (REACH) program. This approach uses an individual's income and Area Deprivation Index (ADI) value associated with their address to create a HEBA score. Scores are then ranked and divided into deciles, which are used to determine adjustments to ACO spending benchmarks, whereas those with the highest deprivation levels get a $30 PMPM adjustment and those with the lowest deprivation levels get a negative $10 PMPM adjustment in 2024.[13]

Data quality

Whether social risk adjustment is occurring within or outside of a risk-scoring algorithm, data is the foundation to develop an objective methodology. The availability and completeness of individual-level data for social risk factors and demographic characteristics like race, ethnicity, and language tend to be much more variable than diagnostic and utilization data. Screening and documentation of health-related social needs (HRSNs) vary considerably across clinical settings and by region; standardized best practices are not yet in place. The use of ICD-10-CM Z-codes to document individual social needs has increased over the last several years, but there is wide geographic variation from state to state and among providers within a state.[14,15] In the datasets used for our analysis, we found that, in both states, between 2.0% and 2.7% of enrollees were coded with at least one Z-code during each year. It is possible that embedding Z-codes into financial adjustments could incentivize more systematic utilization of Z-codes, and it has been observed, in states with managed care, ACO programs, and other value-based payment initiatives, that payer alignment and other types of healthcare delivery transformation that incentivize payers and providers to address health-related social needs tend to have higher rates of documentation using Z-codes.[16]

To improve the quality and completeness of social needs data, policymakers can also look to data sources that live outside of core Medicaid data systems and take steps to integrate datasets from other state agencies or public programs. For example, the Homeless Management Information System (HMIS)[17] is a locally operated repository that collects unduplicated counts of individuals and families experiencing homelessness and their utilization of housing supports and services. Integrating Medicaid data with HMIS can provide a more reliable indicator of homelessness status and enable more nuanced analysis into the relationship between homelessness, health status, and healthcare spending.

13. Navathe, A.S. & Liao, J.M. (September 28, 2023). Embedding Equity in Financial Benchmarks: Changes to the Health Equity Benchmark Adjustment. Health Affairs. Retrieved November 26, 2024, from https://www.healthaffairs.org/content/forefront/embedding-equity-financial-benchmarks-changes-health-equity-benchmark-adjustment#:~:text=A%20key%20equity-oriented%20feature%20in%20ACO%20REACH%20is,benchmarks%20based%20on%20the%20characteristics%20of%20beneficiaries%20served.

14. CMS (September 2021). Utilization of Z Codes for Social Determinants of Health Among Medicare Fee-for-Service Beneficiaries, 2019. Retrieved November 26, 2024, from https://www.cms.gov/files/document/z-codes-data-highlight.pdf.

15. Ubri, P.S. et al. (March 2022). The Role of State Policy in Use of Z Codes to Document Social Need in Medicaid Data. NORC Policy Brief. Retrieved November 26, 2024, from https://norc.org/content/dam/norc-org/pdfs/The%20Role%20of%20State%20Medicaid%20Policy%20in%20Documentation%20of%20SDOH%20in%20Medicaid%20Data_032422.pdf.

16. Ibid.

17. HUD Exchange. HMIS: Homeless Management Information System. Retrieved November 26, 2024, from https://www.hudexchange.info/programs/hmis/.

There is also wide variation in the collection of race/ethnicity data across state Medicaid programs, with only 15 states categorized as having low concern in the collection of these fields in their Transformed Medicaid Statistical Information Systems (T-MSIS).[18] For the analyses conducted in this paper, we applied an imputation methodology to improve the completeness of race/ethnicity data. As state policymakers conduct analytics to identify disparities and adjust payment as an equity strategy, additional consideration and dialogue is warranted to determine whether imputation is appropriate and how it can be applied ethically. States may also want to explore strategies for improving the collection of self-reported race and ethnicity data, such as improved engagement of enrollees and/or enrollment assisters, or enhancement to the enrollment interface.[19] Part of these efforts may include updates to the categories individuals may select to better align with their racial and ethnic identities. Some of these are changes underway, with the U.S. Office of Management and Budget announcing in March 2024 that revisions were to be made to the “Standards for Maintaining, Collecting and Presenting Federal Data on Race and Ethnicity,” intended to better allow individuals to see themselves in the categories they are selecting from.[20] States can embark on these changes earlier rather than waiting for federal updates to go into effect.[21]

While including explicit adjustments to a risk adjustment model based on member race/ethnicity may introduce ethical complexities, having good demographic data enables the exploration of disparities and testing for whether risk adjustment is inadvertently driving resources away from certain groups. Additionally, we can look beyond how social and demographic characteristics predict spending and expand analytic outcomes of interest, such as access to and quality of care. The findings of these analyses can allow states to create objective financial strategies for driving investment to groups with demonstrated barriers to care.

Another option states could consider is including the SDI or other geographic variables in the risk adjustment process. Although our analysis found that these variables are less predictive of utilization patterns than member-specific characteristics (such as race/ethnicity), they may capture some similar effects if the member-level characteristics are excluded from the model, albeit at a less granular level.

Ultimately, the purpose of adjusting risk-based payments to payers and providers based on social risk factors is to mitigate disincentives to serve more complex and hard-to-reach individuals and more equitably resource health system providers who serve these communities. Added dollars for certain populations can lead MCOs and risk-bearing providers to compete to serve complex members, such as by partnering with providers to increase access or investing in member-centered activities like enhanced outreach or in-lieu-of services or value-add benefits to address social needs. As state policymakers consider innovative financing strategies to improve health equity, we recommend that these approaches are complemented by other program requirements to improve quality and outcomes for populations that experience health disparities.

BEYOND RISK ADJUSTMENT

As many state Medicaid agencies have identified addressing health equity as a key priority, states are testing a variety of strategies with the array of Medicaid policy levers available to them.[22] These strategies can include expansions of eligibility through coverage of new populations or continuous enrollment for a longer period of time, coverage of new benefits like community health workers or health-related social needs services, or the inclusion of provisions within managed care contracts to measure and reduce disparities.

---

18. Medicaid.gov (2022). Single Topic Map View: Race and Ethnicity. DQ Atlas. Retrieved November 27, 2024, from https://www.medicaid.gov/dq-atlas/landing/topics/single/map?topic=g3m16&tafVersionId=41.

19. Lukanen, E. & Zylia E. (October 1, 2020). Exploring Strategies to Fill Gaps in Medicaid Race, Ethnicity, and Language Data. State Health and Value Strategies, Princeton University. Retrieved November 27, 2024, from https://www.shvs.org/exploring-strategies-to-fill-gaps-in-medicaid-race-ethnicity-and-language-data/.

20. Pillai, D. & Artiga, S. (April 30, 2024). Revisions to Federal Standards for Collecting and Reporting Data on Race and Ethnicity: What Are They and Why Do They Matter? KFF. Retrieved November 27, 2024, from https://www.kff.org/racial-equity-and-health-policy/issue-brief/revisions-to-federal-standards-for-collecting-and-reporting-data-on-race-and-ethnicity-what-are-they-and-why-do-they-matter/#:~:text=The%20revisions%20include%20using%20a%20single%20combined%20question,collection%20of%20more%20detail%20beyond%20the%20minimum%20categories.

21. Oregon Health Authority (May 7, 2019). Race, Ethnicity, Language and Disability (REALD) Data, Health Equity and CCO 2.0. Retrieved November 27, 2024, from https://olis.oregonlegislature.gov/liz/2019R1/Downloads/CommitteeMeetingDocument/197792.

22. Pillai, A. et al. (July 1, 2024). Medicaid Efforts to Address Racial Health Disparities. KFF. Retrieved November 27, 2024, from https://www.kff.org/medicaid/issue-brief/medicaid-efforts-to-address-racial-health-disparities/.

As states consider the enhancement of risk adjustment methodologies to shift incentives in managed care capitation rate setting, there are opportunities to design managed care requirements that can reinforce financial incentives and create accountabilities for outcomes that complement and justify increases in funding. Managed care strategies can be informed by analytic exercises like those described in this paper to focus on specific subpopulations and outcomes of interest to address health disparities that are specific to the state's unique Medicaid population. Given the findings in our analysis, some examples of managed care policy approaches may include:

- **Quality measurement:** Contracts can include financial incentives such as quality withholds or incentive arrangements tied to the reduction of race/ethnicity disparities in emergency department, inpatient, and primary care utilization. The Kaiser Family Foundation (KFF) recently reported that about one-quarter of states include at least one type of financial incentive for reducing disparities as part of their quality measurement strategies.[23]
- **Other quality initiatives:** States can require MCOs to address health disparities and health-related social needs in the Quality Improvement Plan (QIP), Quality Assessment Performance Improvement (QAPI) program, or Performance Improvement Projects (PIPs).[24] For instance, MCOs may be required to implement a PIP related to screening for adverse childhood experiences (ACEs) and enhanced coordination with child and family systems as an approach to address disparities among individuals with documented problems related to upbringing.
- **Enhanced benefits:** In the absence of Medicaid coverage of health-related social needs services, states can encourage their MCOs to offer in-lieu-of services (ILOS) and value-added benefits (VABs). While coverage of ILOS and VABs are at the option of the health plan, states can provide guidance to encourage uniform service standards and payment approaches.[25] States can explore the use of these enhanced benefits to offer housing navigation, transition, and sustaining services.
- **Community investment:** Some states have implemented managed care requirements to reinvest a percentage of revenue or profit into the communities they serve, or to count community investments as part of the medical loss ratio (MLR) calculation.[26] States can provide MCOs with some guidance on how to use these funds. For example, a state could stipulate that community reinvestment go toward specific rural regions with disparities in inpatient utilization to enhance telehealth capabilities or availability of transportation.
- **Health equity plans:** Many states have opted to require an overall health equity plan from MCOs, which complements specific MCO contract requirements to address health disparities, advance equity, and support HRSNs. This requirement pushes MCOs to look across all aspects of their governance, staffing, and operations to advance equity and create sustainability for equity efforts rather than disparate initiatives, often including engagement of those with lived experience as a foundational element.[27]

No single policy option will address the range of health equity issues present in the Medicaid population. However, states can consider the various policy and program levers to align financial incentives and create a suite of approaches that complement and reinforce each other to advance health equity and address health-related social needs.

23. Pillai, A. et al. (July 1, 2024), Medicaid Efforts to Address Racial Health Disparities, op cit.

24. State Health and Value Strategies (April 2024). Compendium of Medicaid Managed Care Contracting Strategies to Promote Health Equity. Princeton University. Retrieved November 27, 2024, from https://www.shvs.org/wp-content/uploads/2024/04/Compendium-of-Medicaid-Managed-Care-Contracting-Strategies-to-Promote-Health-Equity_April-2024.pdf.

25. Medi-Cal. Transformation of Medi-Cal: Community Supports. Retrieved November 27, 2024, from https://www.dhcs.ca.gov/CalAIM/Documents/DHCS-Medi-Cal-Community-Supports-Supplemental-Fact-Sheet.pdf.

26. Cantor, J. et al. (May 10, 2023). Medicaid Reinvestment Requirements Can Improve Community Health and Equity. Health Affairs. Retrieved November 27, 2024, from https://www.healthaffairs.org/content/forefront/medicaid-reinvestment-requirements-emerging-strategy-improve-community-health-and.

27. Pillai, A. et al. (July 1, 2024), Medicaid Efforts to Address Racial Health Disparities, op cit.

Conclusion

Disparities persist in the U.S. healthcare system, and state Medicaid programs are no exception. We identified a range of social and demographic factors that are associated with notable differences in utilization patterns. These differences cannot be explained fully by traditional risk score algorithms, which rely on claims and enrollment data. Supplementing these algorithms with additional data is likely to improve the predictive accuracy of the risk-scoring models. However, this will not necessarily lead to more equitable allocations of resources, as there are potential pitfalls with policy approaches that focus only on risk-scoring mechanisms, such as reinforcing underutilization that may be driven by lack of access to appropriate healthcare services. States can introduce adjustments within or outside of a risk score to account for disparities and more equitably distribute healthcare resources, informed by analyses like those conducted in this study. Complementary to risk adjustment, states could consider additional managed care requirements and/or other policy approaches like new benefits and eligibility expansions to have a holistic approach to addressing health equity. Further, any state efforts should be paired with a focus on enhancing the quality and completeness of social and demographic data collected by the state to tailor the approach and accurately monitor results.

CONTACT

Anders Larson
anders.larson@milliman.com

Erica Baird
erica.baird@milliman.com

Nazanin Dameshghi
nazanin.dameshghi@milliman.com

Libby Bunzli
libby.bunzli@milliman.com

Spencer Polak
spencer.polak@milliman.com

Appendix A

This section includes technical details related to the model and feature selection, theoretical framework, and overall analytical approach used in the Analysis of Healthcare Utilization Patterns section of this report.

Model selection and analytical approach

The initial analysis involved using descriptive statistics to summarize the distribution of key variables and performing bivariate analyses to examine relationships with healthcare utilization metrics. The data exhibited significant zero inflation across multiple dependent variables, indicating the presence of both structural and sampling zeros.

- **Structural zeros:** Instances where healthcare utilization could not occur due to inherent barriers such as lack of access to healthcare providers. These zeros represent an inaccessible healthcare system for certain individuals.
- **Sampling zeros:** Zeros that arise from individuals within the population who, despite having access or potential need, did not utilize healthcare services during the observation period.

Given the challenges of over-dispersion and excess zeros, we evaluated various statistical models. Although Poisson regression is a common choice for count data due to its simplicity, we determined that approach was unsuitable because it could not handle over-dispersion, where the variance significantly exceeds the mean. We then employed a negative binomial regression (NB) to address over-dispersion by introducing an additional parameter to model the variance independently of the mean. While the NB model offered a significant improvement over the Poisson regression, we explored additional models to better handle the excess zeros present in the dataset. (Nguyen & Wang, 2019; Mihaylova et al., 2011) Specifically, we considered zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB) models to effectively manage the excess zeros. These models are designed to account for both structural zeros and sampling zeros.[28,29]

Based on the performance of these models, we used a combination of ZINB and NB models. Specifically, for prenatal visits, postpartum visits, antianxiety medication counts, and antidepressant medication counts, we found the NB model to be more appropriate. For the other healthcare utilization metrics, the ZINB model provided a better fit. This selection process aligns with recent methodological recommendations for analyzing zero-inflated count data in healthcare utilization studies.[30,31] The Vuong test and Akaike's information criterion (AIC) suggested that the ZINB model provided the most significant improvement over traditional models for count outcomes.[32]

We used the Vuong test, commonly applied in practice for demonstrating the appropriateness of a zero-inflated model compared to its non-zero-inflated counterpart, to compare the ZINB model against the NB model. We also used AIC to provide a measure of the relative quality of multiple statistical models, balancing model fit with complexity.[33]

Variable selection and theoretical framework

Our approach to variable selection and model development was guided by both statistical considerations and theoretical underpinnings, primarily the Andersen healthcare utilization model.[34,35]

28. Lee, K. H., Pedroza, C., Avritscher, E. B. C. et al. (2023). Evaluation of Negative Binomial and Zero-Inflated Negative Binomial Models for the Analysis of Zero-Inflated Count Data: Application to the Telemedicine for Children With Medical Complexity Trial. Trials, 24(1), 613.
29. Neelon, B., O'Malley, A.J., & Smith, V.A. (2016). Modeling Zero-Modified Count and Semicontinuous Data in Health Services Research, Part 1: Background and Overview. Statistics in Medicine, 35(27), 5070-5093.
30. Nguyen, H. T. & Wang, W. (2019). Modeling Count Data for Healthcare Utilization: An Empirical Study of Vietnamese Older People. BMC Public Health, 19, 1234.
31. Mihaylova, B., Briggs, A., O'Hagan, A., & Thompson, S. G. (2011). Review of Statistical Methods for Analysing Healthcare Resources and Costs. Health Economics, 20(8), 897-916.
32. Lee, K. H., Pedroza, C., Avritscher, E. B. C. et al. (2023), op cit.
33. Ibid.
34. Andersen, R. M. (1995). Revisiting the Behavioral Model and Access to Medical Care: Does It matter? Journal of Health and Social Behavior, 36(1), 1-10.
35. Babitsch, B., Gohl, D., & von Lengerke, T. (2012). Re-Revisiting Andersen's Behavioral Model of Health Services Use: A Systematic Review of Studies From 1998–2011. GMS Psycho-Social-Medicine, 9, Doc11.

This model categorizes factors influencing healthcare utilization into three main components: predisposing factors (e.g., age, gender, race, health status), enabling factors (e.g., income, education, access to healthcare services), and need factors (e.g., perceived health conditions and need for services).

To ensure the validity and interpretability of our models, we first tested for multicollinearity using variance inflation factors (VIFs) and removed variables that were identified as highly correlated. This step helped to improve model performance and ensure that the remaining variables provided unique contributions to the model.

We adopted a two-stage modeling approach to ensure a comprehensive and theoretically grounded analysis. Initially, we fit a NB regression model to identify significant predictors of healthcare utilization. Variables showing statistical significance in the NB model ($p < 0.05$) were included in the count component of the ZINB model, allowing a focus on the most relevant predictors while maintaining model parsimony.[36]

For the zero-inflated component of our ZINB model, we employed a targeted variable selection approach based on the Andersen model and prior research in healthcare utilization. Key predisposing and enabling factors influencing the likelihood of non-utilization were identified, including demographic variables (e.g., race), socioeconomic indicators (e.g., SDI components, including the percentage of the population below 100% FPL), and access-related factors (e.g., rural/urban residence, food shortage, and the percentage of households without a vehicle). Additionally, we included homelessness status, mental health disorder status, and substance use disorder status as key variables within the Andersen framework, given their critical role in healthcare utilization patterns among vulnerable populations.[37,38] Lastly, for the models predicting inpatient admissions and emergency department visits, we included the appropriate MARA service category risk score in the zero-inflated component. They are low-frequency events and we found that, if the MARA risk score was not included in the zero-inflated component, the impact of other features was overstated.[39]

Statistical criteria, including significance tests and AIC, complemented this theoretically driven selection to refine the final set of variables in the zero-inflated component. Integrating the Andersen model into the variable selection process ensured a comprehensive analysis of the complex factors influencing healthcare utilization patterns. This approach allowed for a systematic examination of how predisposing, enabling, and need factors contribute to both structural and sampling zeros in the data. The resulting model provides a nuanced understanding of healthcare utilization patterns, accounting for the inherent complexities in healthcare data and the various factors influencing both utilization and non-utilization of healthcare services.

This comprehensive methodology, grounded in both statistical rigor and theoretical understanding, allows for a robust analysis of healthcare utilization patterns, accounting for the complexities inherent in healthcare data and the various factors that influence both utilization and non-utilization of healthcare services. We conducted all analyses using R software (version 4.2.1), with the “glmmTMB” package (version 1.1.8) for ZINB modeling.

With the exception of the models predicting prenatal and postpartum visits, the final models employed in this analysis had R-squared values ranging from 21.3% to 58.3%. R-squared values were highest for models predicting pharmacy utilization and lowest for models predicting inpatient admissions and emergency department visits. The R-squared values were notably lower for the prenatal and postpartum visit models (1.2% and 2.2%, respectively), indicating that the variables considered in this analysis do not explain much of the variation for these visits.

36. Nguyen, H. T. & Wang, W. (2019), Modeling Count Data, op cit.

37. Gelberg, L., Andersen, R. M., & Leake, B. D. (2000). The Behavioral Model for Vulnerable Populations: Application to Medical Care Use and Outcomes for Homeless People. Health Services Research, 34(6), 1273-1302.

38. Phillips, K. A., Morrison, K. R., Andersen, R., & Aday, L. A. (1998). Understanding the Context of Healthcare Utilization: Assessing Environmental and Provider-Related Variables in the Behavioral Model of Utilization. Health Services Research, 33(3 Pt 1), 571-596.

39. This was particularly notable for the mental health and SUD flags. When the MARA score was not included in the zero-inflation component, the model indicated that individuals with at least one mental health or SUD condition had roughly two to three times as many inpatient admissions. When the MARA score was included in zero-inflation component, these relationships became much weaker, suggesting that the MARA risk score was already capturing most of these effects.

Appendix B

Code set for identifying primary care visits - Taxonomy codes included

NATIONAL UNIFORM CLAIM COMMITTEE (NUCC) TAXONOMY CODE	NUCC CLASSIFICATION	NUCC SPECIALIZATION
207RG0300X	Internal Medicine	Geriatric Medicine
207QG0300X	Family Medicine	Geriatric Medicine
2083X0100X	Preventive Medicine	Occupational Medicine
207QH0002X	Family Medicine	Hospice and Palliative Medicine
207Q00000X	Family Medicine	Undefined
207QA0401X	Family Medicine	Addiction Medicine
207QA0505X	Family Medicine	Adult Medicine
207QB0002X	Family Medicine	Obesity Medicine
207QS0010X	Family Medicine	Sports Medicine
207QS1201X	Family Medicine	Sleep Medicine
207RH0002X	Internal Medicine	Hospice and Palliative Medicine
208D00000X	General Practice	General Practice (Undefined)
2083A0300X	Preventive Medicine	Addiction Medicine
2083A0100X	Preventive Medicine	Aerospace Medicine
2083P0500X	Preventive Medicine	Preventive Medicine/Occupational Environmental Medicine
2083P0901X	Preventive Medicine	Public Health & General Preventive Medicine
2083S0010X	Preventive Medicine	Sports Medicine
2083T0002X	Preventive Medicine	Medical Toxicology
207QA0000X	Family Medicine	Adolescent Medicine
207RA0000X	Internal Medicine	Adolescent Medicine
208000000X	Pediatrics	Undefined
2080A0000X	Pediatrics	Adolescent Medicine
2080H0002X	Pediatrics	Hospice and Palliative Medicine
2080N0001X	Pediatrics	Neonatal-Perinatal Medicine
2080C0008X	Pediatrics	Child Abuse
2080B0002X	Pediatrics	Obesity Medicine
2080S0012X	Pediatrics	Sleep Medicine
2080S0010X	Pediatrics	Sports Medicine
363LA2200X	Nurse Practitioner	Adult Health Nurse Practitioner
363LF0000X	Nurse Practitioner	Family Nurse Practitioner
363LP2300X	Nurse Practitioner	Primary Care Nurse Practitioner
363A00000X	Physician Assistant	Physician Assistant
363AM0700X	Physician Assistant	Medical Physician Assistant
261QF0400X	Clinic/Center	Federally Qualified Health Center
261QR1300X	Clinic/Center	Rural Health Clinic

Code set for identifying primary care visits - CPT codes included

CPT CODE	DESCRIPTION
99201	New patient office visit with a problem-focused history and exam, and straightforward decision making, typically involving 10 minutes face-to-face.
99202	New patient office visit with medically appropriate history/exam and straightforward decision making, requiring at least 15 minutes of total time.
99203	New patient office visit with medically appropriate history/exam and low-level decision making, requiring at least 30 minutes of total time.
99204	New patient office visit with medically appropriate history/exam and moderate decision making, requiring at least 45 minutes of total time.
99205	New patient office visit with medically appropriate history/exam and high-level decision making, requiring at least 60 minutes of total time.
99211	Established patient office visit that may not need a physician's presence.
99212	Established patient office visit with medically appropriate history/exam and straightforward decision making, requiring at least 10 minutes of total time.
99213	Established patient office visit with medically appropriate history/exam and low-level decision making, requiring at least 20 minutes of total time.
99214	Established patient office visit with medically appropriate history/exam and moderate decision making, requiring at least 30 minutes of total time.
99215	Established patient office visit with medically appropriate history/exam and high-level decision making, requiring at least 40 minutes of total time.
99241	Office consultation for a new or established patient with problem-focused history/exam and straightforward decision making, typically involving 15 minutes face-to-face.
99242	Office consultation for a new or established patient with medically appropriate history/exam and straightforward decision making, requiring at least 20 minutes of total time.
99243	Office consultation for a new or established patient with medically appropriate history/exam and low-level decision making, requiring at least 30 minutes of total time.
99244	Office consultation for a new or established patient with medically appropriate history/exam and moderate decision making, requiring at least 40 minutes of total time.
99245	Office consultation for a new or established patient with medically appropriate history/exam and high-level decision making, requiring at least 55 minutes of total time.
99441	Telephone evaluation for an established patient with 5-10 minutes of medical discussion, not related to a recent or upcoming E/M service.
99442	Telephone evaluation for an established patient with 11-20 minutes of medical discussion, not related to a recent or upcoming E/M service.
99443	Telephone evaluation for an established patient with 21-30 minutes of medical discussion, not related to a recent or upcoming E/M service.

Appendix C

List of rate cells included

POPULATION	RATE CELL	STATES INCLUDED
TANF (Low-Income Families and Children)	TANF 1-6 Years Male/Female	2
TANF (Low-Income Families and Children)	TANF 7-13 Years Male/Female	2
TANF (Low-Income Families and Children)	TANF 14-18 Years Female	2
TANF (Low-Income Families and Children)	TANF 14-18 Years Male	2
TANF (Low-Income Families and Children)	TANF 19-44 Years Female	2
TANF (Low-Income Families and Children)	TANF 19-44 Years Male	2
TANF (Low-Income Families and Children)	TANF 45-64 Male/Female	2
SSI (Disabled)	SSI 1-18 Years Male/Female	2
SSI (Disabled)	SSI 19-64 Years Male/Female	2
Foster Care	Foster Care Children Male/Female	2
Medicaid Expansion	Expansion 19-24 Years Female	1
Medicaid Expansion	Expansion 19-24 Years Male	1
Medicaid Expansion	Expansion 25-34 Years Female	1
Medicaid Expansion	Expansion 25-34 Years Male	1
Medicaid Expansion	Expansion 35-44 Years Female	1
Medicaid Expansion	Expansion 35-44 Years Male	1
Medicaid Expansion	Expansion 45-54 Years Female	1
Medicaid Expansion	Expansion 45-54 Years Male	1
Medicaid Expansion	Expansion 55-64 Years Female	1
Medicaid Expansion	Expansion 55-64 Years Male	1