MILLIMAN REPORT

Testing Milliman Advanced Risk Adjuster models for racial bias

Medicare model results

Erica Rode, PhD, FSA, MAAA
Hans Leida, PhD, FSA, MAAA

Milliman

Careful consideration when it comes to modeling choices, variable selection, and program design can help reduce the potential for bias in risk adjustment.

Recently there has been growing concern that algorithms used for care management may exhibit racial bias. Where it exists, this bias arises due to the ways certain algorithms use past experience to predict future costs, which may be used to identify individuals for targeted care management program interventions. We take this concern seriously. We are committed to understanding and eliminating any potential for perpetuating or worsening racial inequities within the healthcare system. Mindful of the role Milliman Advanced Risk Adjusters™ (MARA™) plays in assessing health risks, we are investigating our models for any indication of this bias. Using the data on Medicare fee-for-service (FFS) beneficiaries from the 5% sample of the Enrollment DataBase (EDB) and National Claims History Standard Analytical Files (SAFs) released by the Centers for Medicare and Medicaid Services (CMS),[1] we tested two prospective diagnosis-based MARA models intended for use on Medicare populations, and found no indication of racial bias using the definition described below.

Racial bias in risk adjustment

HOW IS RACIAL BIAS DEFINED?

A recent study published in *Science* found racial bias in a widely used algorithm, using the following definition: "At a given risk score, Black patients are considerably sicker than White patients."[2] In other words, the algorithmic bias results in lower risk scores for Black people who are not in reality healthier than white people. In that study, the classification of a patient as "sicker" was based on the prevalence and severity of certain chronic conditions, as determined by various electronic health record (EHR) measures. The study classifies an algorithm as racially biased if members of one race cohort are more likely to be prioritized for care management than another race cohort, at the same level of health, based on having a higher risk score.

WHAT CAUSES THIS RACIAL BIAS?

The misalignment between health status and risk score identified in the *Science* article is likely attributable to a number of causes, both societal and analytical, including:

1. Lower-income people are less likely to have the resources necessary to seek care and may face barriers to obtaining care, such as lack of transportation or an inability to take time off work. As a result, conditions may not be documented in medical claims and therefore do not contribute to a risk score. This will disproportionately affect Black people to the degree there is a correlation between income and race for a particular study population.
2. The majority of risk-scoring algorithms are designed to predict healthcare costs rather than illness burden, and the two are not always in alignment.
3. Furthermore, many algorithms use information on procedures performed in the past or on prior costs for the individual in modeling. Yet these variables do not always result in higher performance and may also unintentionally lead to bias to the degree cost differs among groups for reasons other than morbidity.
4. Research has shown that the quality of care received by Black people is lower than that received by white people.[3] Black people also tend to have less trust in the healthcare system and therefore are less likely to visit a physician and have diagnoses coded in their medical records.[4] This contributes to reluctance to visit and communicate openly with physicians, as well as worse clinical outcomes for Black people.

[1] CMS. National Claims History (NCH) Standard Analytical Files. Retrieved September 3, 2020, from https://aspe.hhs.gov/centers-medicare-medicaid-services (scroll down to NCH SAF).

[2] Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (October 25, 2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science Magazine. Retrieved September 3, 2020, from https://science.sciencemag.org/content/366/6464/447.

[3] AHRQ (September 2019). 2018 National Healthcare Quality and Disparities Report. Retrieved September 3, 2020, from https://www.ahrq.gov/research/findings/nhqrdr/nhqdr18/index.html.

[4] Armstrong, K., Ravenell, K. L., McMurphy, S., & Putt, M. (July 2007). Racial/ethnic differences in physician distrust in the United States. Am J Public Health. Retrieved September 3, 2020, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1913079/.

Testing MARA for bias

WHAT IS MARA?

MARA is a suite of risk adjustment models developed by healthcare actuaries and other professionals at Milliman. This paper focuses on the prospective Medicare DxXPLN and DxOPTml models. These models use 12 months of administrative medical claim data and demographics to predict relative healthcare resource use for an individual for the 12-month period immediately following the claim period. They are calibrated specifically for a Medicare population, and predict total relative per member per month (PMPM) medical and prescription drug allowed costs. These models were chosen for this assessment of bias because they are appropriate for the available data, which is a Medicare population and contains only medical claims. The DxXPLN models use regularized regression, while the DxOPTml models use advanced machine learning algorithms to generate predictions. The MARA models use numerous details from claim data, but do not use prior cost and make limited use of procedure codes. These data elements were intentionally not used to avoid inflating risk scores in cases where there is overutilization because of poor care rather than morbidity.

MARA is used by healthcare providers, health plans, care managers, and others for a variety of purposes, including pricing, measuring outcomes, and targeted interventions. In addition to a total risk score for each individual, MARA outputs a wealth of other information, including a complete clinical profile and risk scores by service category to help users identify risk drivers and focus care management efforts.

STUDY DESIGN

MARA was tested for racial bias using metrics similar to those described by the researchers in the aforementioned study. A sample of individuals was scored, and individuals were stratified by risk score. Scores were normalized to an average of 1.00 over the population being studied. Then the number of chronic conditions in each race cohort in the subsequent 12-month period was compared to see if the health status, as measured by the prevalence of chronic conditions, was materially different for Black people and white people at similar risk score levels. Due to limitations in the available data, we were unable to perform other analyses presented in the *Science* article, such as the comparison of condition severity based on EHR measures.

The data used is from the 5% sample of Medicare beneficiary data from 2016 and 2017 as described above. This data was not used in calibration or development of any the MARA models tested in this analysis. It was chosen for this analysis due to the presence of race identifiers in the data, as collected by the Social Security Administration. To the extent there is bias in the data regarding coding of race, this could cause bias in our sample. Medical claims and eligibility data from calendar year 2016 were used to assign risk scores predicting utilization for calendar year 2017. Chronic conditions were identified in calendar year 2017 using MARA's clinical contribution output, which categorizes all coded conditions for each individual and identifies those considered chronic, based on guidance released by the Agency for Healthcare Research and Quality (AHRQ).[5]

We excluded the following people from the analysis:

- Beneficiaries who did not have eligibility in all 12 months of 2017
- Beneficiaries whose race values were not consistent over the study time period
- Beneficiaries with invalid birthdates
- Beneficiaries who were not enrolled in both Medicare Part A and Part B fee-for-service coverage for the study time period
- Beneficiaries in hospice or with end-stage renal disease (ESRD) in 2016

The original sample included 1,246,197 white people and 135,247 Black people after these exclusions. There were also 47,431 people with an "unknown" or "other" race, and a combined total of 65,387 identified as Asian, Hispanic, or North American Native. The data volume for these other racial cohorts was not considered sufficient for use in this analysis, and therefore it was excluded. The table in Figure 1 shows descriptive statistics for the original sample.

FIGURE 1: DESCRIPTIVE STATISTICS OF ORIGINAL SAMPLE

	WHITE	BLACK
LIVES	1,246,197	135,247
AVERAGE AGE	72	66
FEMALE PERCENTAGE	55%	56%
AVERAGE DxXPLN RISK SCORE	0.99	1.08
AVERAGE DxOPTml RISK SCORE	0.99	1.13
AVERAGE PMPM 2016 ALLOWED CLAIM COST	$794	$831
AVERAGE PMPM 2017 ALLOWED CLAIM COST	$929	$978

[5] Healthcare Cost and Utilization Project (May 2016). Chronic Condition Indicator. AHRQ. Retrieved September 3, 2020, from https://www.hcup-us.ahrq.gov/toolssoftware/chronic/chronic.jsp.

As Figure 1 shows, the average age was significantly different for each race cohort. Age plays a material role in prospective risk scores, as well as patient morbidity, so that two individuals of different ages with the same risk score may not necessarily have similar health statuses. This is particularly true in Medicare, where disability or ESRD is required for eligibility before age 65.

To ensure that results were not skewed by these differences, we completed our analysis on a subset of the white population sample that was randomly chosen to achieve a similar distribution of ages between the two race cohorts (that is, we took a stratified random sample). Results for the original sample are included in the appendix. We also performed sensitivity testing of results by performing the analysis using two additional stratified random samples of this type. In this report, we will refer to these three random subsets as stratified samples A, B, and C. The table in Figure 2 shows descriptive statistics for stratified sample A. Statistics for samples B and C are shown in the appendix.

FIGURE 2: DESCRIPTIVE STATISTICS OF STRATIFIED SAMPLE A

	WHITE	BLACK
MEMBERS	135,247	135,247
AVERAGE AGE	66	66
FEMALE PERCENTAGE	53%	56%
AVERAGE DxXPLN RISK SCORE	0.97	1.03
AVERAGE DxOPTml RISK SCORE	0.94	1.06
AVERAGE PMPM 2016 ALLOWED CLAIM COST	$794	$831
AVERAGE PMPM 2017 ALLOWED CLAIM COST	$915	$978

In order to measure the level of statistical significance of the differences seen above, we also used bootstrapping techniques to develop confidence intervals and ranges of plausible outcomes.

RESULTS

Using the metrics described above, the MARA Medicare DxXPLN and DxOPTml models show no evidence of racial bias on any of the population samples we tested. Below we show some key metrics for the population after sampling to a similar age distribution between the race cohorts.

Figure 3 graphs the chronic condition counts per individual by risk score percentile for each model for stratified sample A. Results for the full population, as well as results of sensitivity testing by using the different random samples B and C of the white population, are shown in the appendix.

FIGURE 3: CHRONIC CONDITION COUNT PER INDIVIDUAL BY RISK SCORE PERCENTILE, SAMPLE A, DxOPTml

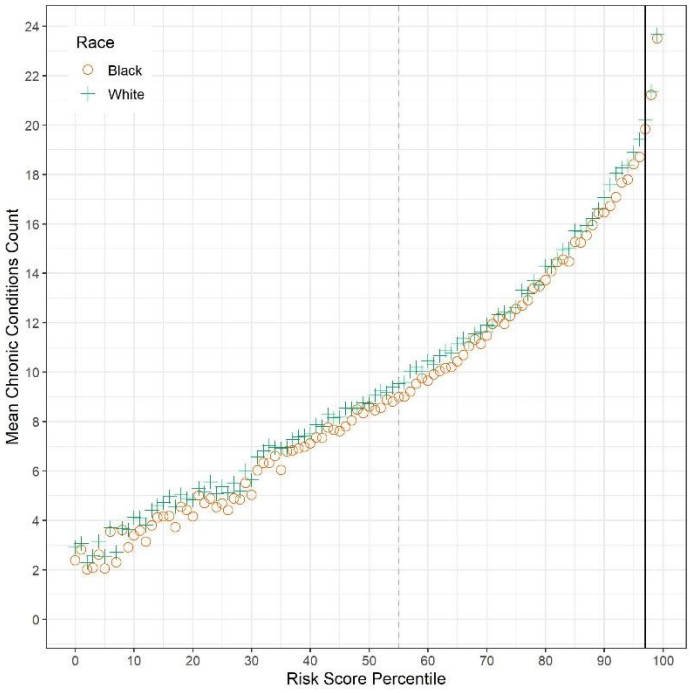


FIGURE 4: AVERAGE CHRONIC CONDITION COUNTS PER INDIVIDUAL, SAMPLE A, DxOPTml

	WHITE	BLACK	DIFFERENCE*
ALL MEMBERS	9.3	9.4	0.5%
TOP 1%	23.7	23.5	-0.6%
TOP 3%	21.7	21.5	-0.8%
TOP 5%	20.7	20.4	-1.4%

* Calculated as (Black – White) / White.

The negative difference amounts show that the Black cohort has slightly fewer chronic conditions per individual, on average, at each risk score cutoff. This is the opposite of the relationship discussed in the *Science* article cited above, suggesting that, if anything, Black people have higher risk scores than white people at the same levels of health. However, the difference in chronic condition prevalence between the two race cohorts for these samples is quite small on a percentage basis, and our sensitivity testing (included in the appendix) did not show directionally consistent results. Furthermore, it should be noted that there are approximately 2,700 people within each risk score percentile, and we see variation in the average risk score at that level, meaning that not all people within the same risk score percentile are being assigned the same morbidity level by MARA.

FIGURE 5: RISK SCORE AVERAGES, SAMPLE A, DxOPTml

	WHITE	BLACK	DIFFERENCE*
ALL MEMBERS	0.94	1.06	11.9%
TOP 1%	6.72	6.84	1.7%
TOP 3%	4.97	5.05	1.6%
TOP 5%	4.22	4.31	2.1%

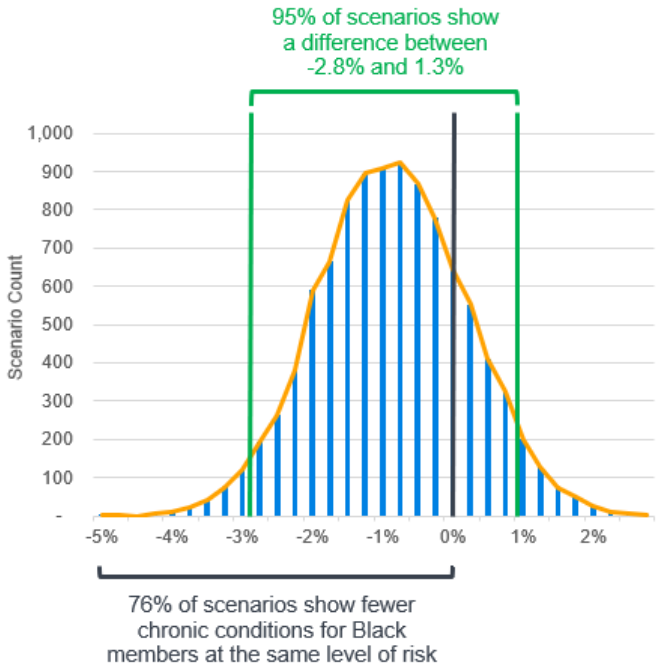
* Calculated as (Black – White) / White.

Note that the Black cohort has a consistently higher risk score than the white cohort at every percentile cutoff, even those where the chronic condition count is lower. It should also be noted that, when considering the age-matched sample, the difference in risk score is greater than the difference in cost. During sensitivity testing, we observed similar results in samples B and C as well.

Looking at the differences in Figures 4 and 5, it is important to bear in mind that the count of chronic conditions measure in Figure 4 does not take into account the mix of chronic conditions underlying each cohort. Differences in the clinical profile of each cohort may also be relevant to understanding their relative health status. Risk scores like those in Figure 5 are intended to reflect the impact of such differences in clinical profile.

To further understand how much variation in risk score could be expected due to random fluctuation, we used a bootstrapping approach. We sampled the population in sample A with replacement to produce 10,000 random samples of the same size as the original, and compared the outcomes for each sample. This gives us insight into the level of random fluctuation by analyzing outcomes if the population distribution changed. In all 10,000 scenarios, the difference in the count of chronic conditions between Black and white people in the top 3% of risk scores never exceeded 5%. The graph in Figure 6 shows the distribution of outcomes, which illustrates that neither race cohort consistently showed a higher prevalence of chronic conditions at the same level of risk compared to the other race cohort. Based on this, we do not have evidence to support the hypothesis that the MARA models are biased against either cohort.

FIGURE 6: DISTRIBUTION OF CHRONIC CONDITION PERCENTAGE DIFFERENCES (BLACK – WHITE) FOR TOP 3% OF RISK SCORES, SAMPLE A, DxOPTml



WHAT MAKES MARA DIFFERENT?

Why did MARA not exhibit the same type of racial bias as the model discussed in the *Science* article cited above? When designing the MARA models we took careful consideration to avoid rewarding excessive utilization or factoring treatment patterns into the risk score. Instead, MARA was designed to reflect expected healthcare resource utilization based on an individual's morbidity level. We hypothesize that the following factors contribute to the lack of observed bias:

1. Not using prior cost as a predictor in MARA risk adjustment models avoids increasing risk scores based on past expenditures, which may be low among low-income populations. To the degree there is a relationship between race and income, the use of prior cost could introduce bias.
2. MARA makes limited use of procedure codes, which helps reduce the impact of variation in treatment for the same conditions. This may contribute toward reducing bias by removing the impact of high-cost or elective procedures for individuals with similar conditions but different income levels, to the extent there is a correlation between treatment for a given condition and patient income level.

3. MARA also considers the venue in which a diagnosis was recorded to recognize that people receiving treatment in emergency departments and inpatient stays may have higher risks of poor health outcomes or acute events in the projection period than those receiving care through physician office visits or other planned avenues of care. This also promotes recognition of the cultural differences in health-seeking behavior.

We also analyzed performance of the models on the age-matched samples, using relative risk score to predict relative allowed costs in the projection period, and found that:

1. Measured on sample A, the DxOPTml model slightly overpredicted costs for the Black cohort (by 2%) and underpredicted costs for the white cohort (by 3%). The DxXPLN model did not overpredict or underpredict either cohort to a material degree (predictions were within 0.2%).
2. Both the DxOPTml and DxXPLN models had a comparable R^2 on both cohorts. R^2 is a measure of how much variation in scores is explained by the model and is very sensitive to outliers.
3. Both models exhibited a higher mean absolute prediction error (MAPE) on the Black cohort compared to the white cohort. This indicates that the model does not predict costs as accurately for the Black cohort as the white cohort.

Other considerations

CARE MANAGEMENT CONSIDERATIONS

Although risk scores can be a useful tool in prioritizing people for care management, the scores themselves are just one element in a comprehensive candidate identification program. There are various other aspects of a patient's health that should be considered, such as:

1. Complexity of healthcare needs. Is the high risk score driven by one very severe condition, such as hemophilia, or does the individual have multiple comorbid conditions? Mental health and substance use disorders in particular can exacerbate, and be exacerbated by, other chronic conditions.
2. The extent to which a patient's conditions are already being managed. Is the patient currently taking medication for his or her conditions? Does the clinical profile show a high likelihood of an emergency department visit in the next 12 months? Social determinants, such as housing instability or food insecurity, also play an important role in an individual's ability to manage conditions without supports such as care management interventions, and should be considered.

MARA outputs a wealth of additional information about individuals, including:

1. A complete, categorized listing of all conditions that a person has in the exposure period.
2. Chronic and complexity indicators for each condition.
3. Service category risk scores. Inpatient and emergency department scores are highly correlated with prospective inpatient hospital admits and emergency department visits, respectively.

OTHER BIAS DEFINITIONS AND OTHER USE CASES

The definition of bias used in the *Science* article and that we have explored in this paper is a reasonable one to use and provides useful insights, but one should consider some alternatives as well. In a model used for care management, instead of comparing the health status of individuals at the same level of risk score, we might explore the rate at which individuals of the same level of health are identified for care management using risk scores to determine if there are sources of bias in the implementation of the care management program, outside of the risk score itself. Further, this analysis does not capture the impact of diagnoses that are never recorded for a member. To the extent that undiagnosed chronic conditions are more prevalent in one race cohort than another, this could be another source of bias.

Any definition of bias should be considered in the context of how the risk-scoring model is being used. Risk adjustment models are used for many purposes besides care management, where the impact of racial differences could have many different consequences. For example, risk scores are often used for payment purposes, such as Patient Protection and Affordable Care Act (ACA) risk adjustment transfers, setting capitation rates for Medicaid managed care organizations, and determining payments to Medicare Advantage and Part D organizations. Each of these presents unique challenges and consequences that should be considered when choosing and evaluating a risk adjustment methodology.

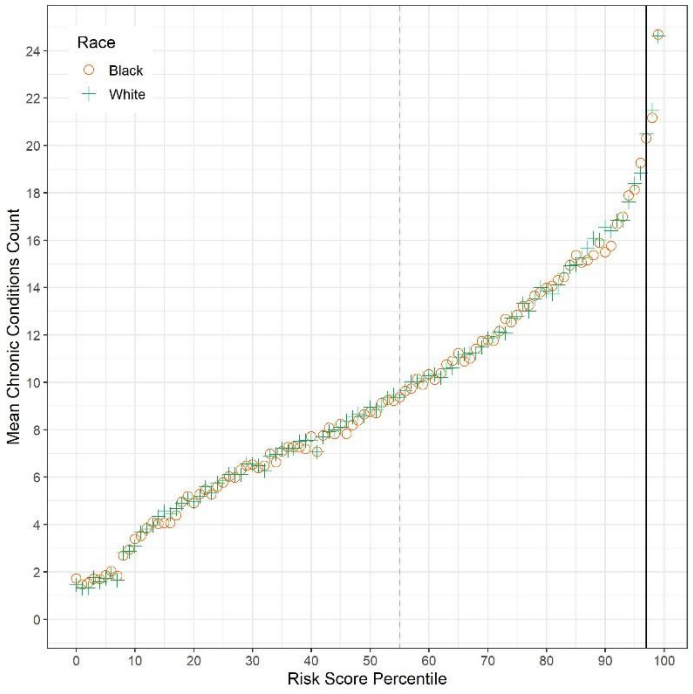
Appendix: Additional results

DxXPLN RESULTS ON SAMPLE A

Here we provide detailed results for the DxXPLN model on the same age-matched sample as those presented for the DxOPTml model in the body of the report.

First, we looked at a comparison chronic condition count by risk score percentile and, similar to the DxOPTml model, we do not see a material difference in condition count between the two race cohorts.

FIGURE 7: CHRONIC CONDITION COUNT BY RISK SCORE PERCENTILE, SAMPLE A, DxXPLN



The table in Figure 8 shows that, for this sample, there is even less difference between the condition counts than in the DxOPTml model.

FIGURE 8: CHRONIC CONDITION COUNTS, SAMPLE A, DxXPLN

	WHITE	BLACK	DIFFERENCE*
ALL MEMBERS	9.3	9.4	0.5%
TOP 1%	24.6	24.7	0.2%
TOP 3%	22.1	22.1	-0.1%
TOP 5%	20.7	20.8	0.3%

* Calculated as (Black – White) / White.

Furthermore, the DxXPLN scores are also higher for Black people than white people at the same risk score percentile, and by an even wider margin than the DxOPTml model for the top 5% and above.

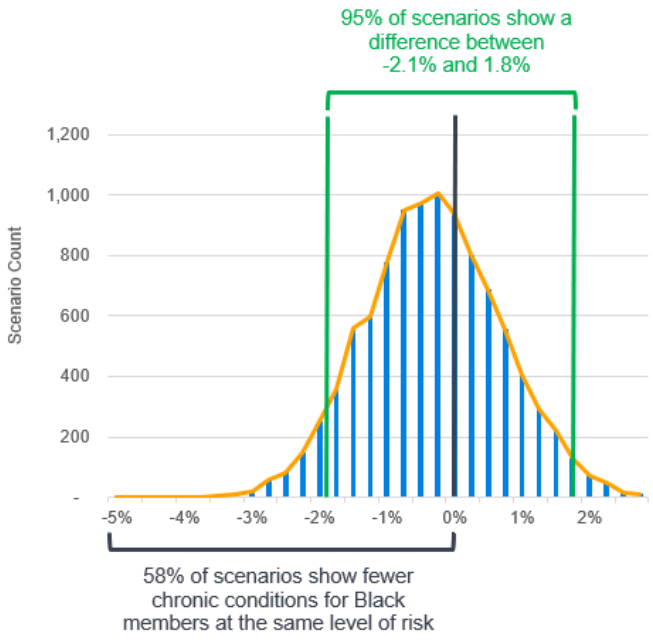
FIGURE 9: RISK SCORE AVERAGES, SAMPLE A, DxXPLN

	WHITE	BLACK	DIFFERENCE*
ALL MEMBERS	0.97	1.03	6.0%
TOP 1%	6.49	6.62	2.0%
TOP 3%	4.93	5.05	2.6%
TOP 5%	4.25	4.34	2.2%

* Calculated as (Black – White) / White.

The graph in Figure 10 shows a similar 95% confidence interval for the difference in chronic condition counts across scenarios, although the results are more balanced than those for the DxOPTml model shown in Figure 6 above.

FIGURE 10: DISTRIBUTION OF CHRONIC CONDITION PREVALENCE DIFFERENCES (BLACK – WHITE) FOR TOP 3% OF RISK SCORES, SAMPLE A, DxXPLN



SENSITIVITY TESTING RESULTS

This section contains the results of our sensitivity testing analysis. Here we repeated the analysis shown in the body of the report for DxOPTml and the prior section for DxXPLN, but using two alternate age-matched samples of white people.

The table in Figure 11 shows descriptive statistics for sample B. Note that the Black cohort is the same in both exhibits, and only the white cohort changes. As shown, sample B is very similar to sample A, still exhibiting lower costs and risk scores for the white cohort overall.

FIGURE 11: DESCRIPTIVE STATISTICS OF STRATIFIED SAMPLE B

	WHITE	BLACK
MEMBERS	135,247	135,247
AVERAGE AGE	66	66
FEMALE PERCENTAGE	53%	56%
AVERAGE DxXPLN RISK SCORE	0.97	1.03
AVERAGE DxOPTml RISK SCORE	0.95	1.06
AVERAGE PMPM 2016 COST	$796	$831
AVERAGE PMPM 2017 COST	$922	$978

Figures 12 and 13 show the distributions of chronic conditions by risk score percentile for sample B, which are similar to sample A. The most notable difference is at the 99th percentile, where the white cohort shows higher chronic condition counts than the Black cohort for both models.

FIGURE 12: CHRONIC CONDITION COUNT BY RISK SCORE PERCENTILE, SAMPLE B, DxOPTml

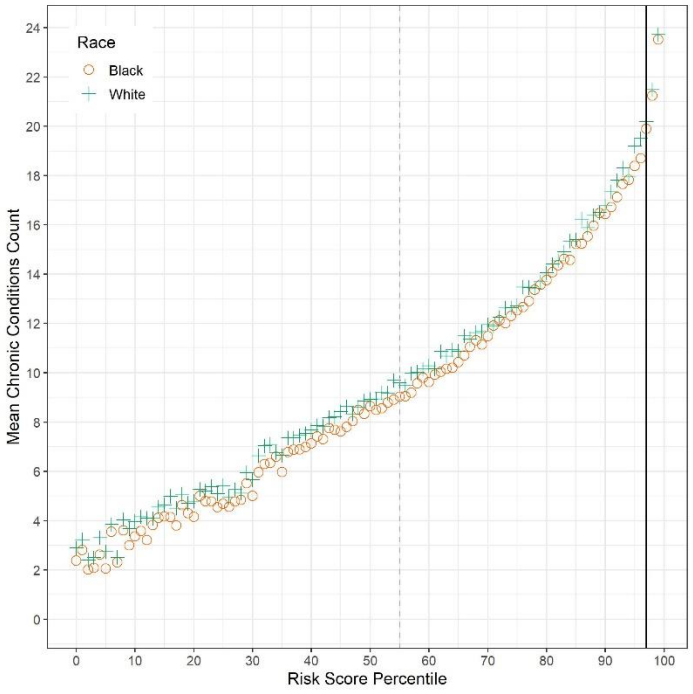
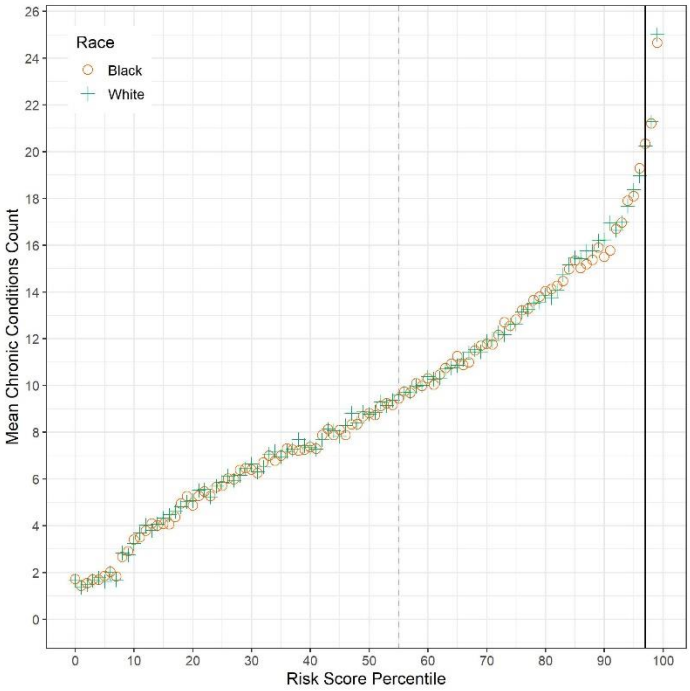


FIGURE 13: CHRONIC CONDITION COUNT BY RISK SCORE PERCENTILE, SAMPLE B, DxXPLN



The tables in Figures 14 and 15 show the chronic condition counts for sample B at various risk score percentiles. Negative differences correspond to lower counts of chronic conditions for the Black cohort, while positive differences correspond to higher counts of chronic conditions. Note that, contrary to sample A, the difference is not directionally consistent across all samples and models.

FIGURE 14: CHRONIC CONDITION COUNTS, SAMPLE B, DxOPTml

	WHITE	BLACK	DIFFERENCE*
ALL MEMBERS	9.3	9.4	0.2%
TOP 1%	23.7	23.5	-0.8%
TOP 3%	21.8	21.6	-1.0%
TOP 5%	20.8	20.4	-2.0%

* Calculated as (Black – White) / White.

FIGURE 15: CHRONIC CONDITION COUNTS, SAMPLE B, DxXPLN

	WHITE	BLACK	DIFFERENCE*
ALL MEMBERS	9.3	9.4	0.2%
TOP 1%	25.0	24.7	-1.4%
TOP 3%	22.1	22.1	0.2%
TOP 5%	20.7	20.8	0.4%

* Calculated as (Black – White) / White.

Below we repeat our analysis on a third random sample of the white population. Once again, we see a similar profile for the white sample, with consistently lower costs and risk scores than the Black cohort.

FIGURE 16: DESCRIPTIVE STATISTICS OF STRATIFIED SAMPLE C

	WHITE	BLACK
MEMBERS	135,247	135,247
AVERAGE AGE	66	66
FEMALE PERCENTAGE	53%	56%
AVERAGE DxXPLN RISK SCORE	0.97	1.03
AVERAGE DxOPTml RISK SCORE	0.95	1.06
AVERAGE PMPM 2016 COST	$797	$831
AVERAGE PMPM 2017 COST	$919	$978

Figures 17 and 18 show a distribution of chronic conditions by risk score percentile for each race cohort that is similar to sample A.

FIGURE 17: CHRONIC CONDITION COUNT BY RISK SCORE PERCENTILE, SAMPLE C, DxOPTml

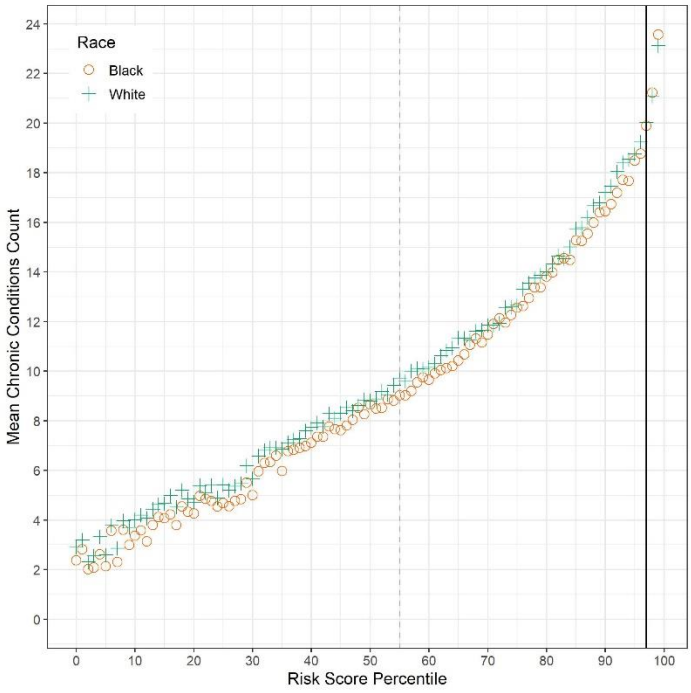
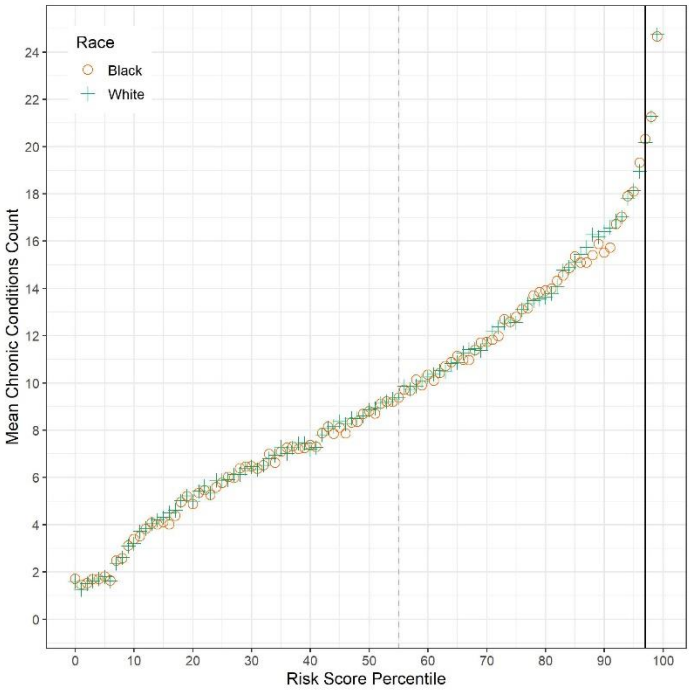


FIGURE 18: CHRONIC CONDITION COUNT BY RISK SCORE PERCENTILE, SAMPLE C, DxXPLN



The tables in Figures 19 and 20 show that the chronic condition count for the white cohort in this sample is generally lower than the count for the Black cohort at each risk score percentile.

FIGURE 19: CHRONIC CONDITION COUNTS, SAMPLE C, DxOPTml

	WHITE	BLACK	DIFFERENCE*
ALL MEMBERS	9.3	9.4	0.3%
TOP 1%	23.1	23.6	1.9%
TOP 3%	21.4	21.6	0.8%
TOP 5%	20.4	20.4	0.0%

* Calculated as (Black – White) / White.

FIGURE 20: CHRONIC CONDITION COUNTS, SAMPLE C, DxXPLN

	WHITE	BLACK	DIFFERENCE*
ALL MEMBERS	9.3	9.4	0.3%
TOP 1%	24.7	24.7	-0.3%
TOP 3%	22.0	22.1	0.4%
TOP 5%	20.6	20.8	0.9%

* Calculated as (Black – White) / White.

Overall, our sensitivity testing does not support a conclusion that either model is consistently biased in favor of either race cohort using this definition. Any differences observed are within a level that could reasonably be attributed to random fluctuation in health status and risk scores rather than actual differences in the cohorts.

RESULTS ON FULL SAMPLE

This section shows results for the DxXPLN and DxOPTml models on the full 5% sample after removing the beneficiaries described in the section on study design. As we noted above, individuals in the Black population are consistently younger, indicating a higher rate of disabled individuals than in the white population. Risk scores in MARA Medicare for disabled people under age 65 are calculated using a different model than is used for people age 65 and over to recognize the inherent differences in these two populations.

FIGURE 21: DESCRIPTIVE STATISTICS OF FULL 5% SAMPLE

	WHITE	BLACK
MEMBERS	1,246,197	135,247
AVERAGE AGE	72	66
FEMALE PERCENTAGE	55%	56%
AVERAGE DxXPLN RISK SCORE	0.99	1.08
AVERAGE DxOPTml RISK SCORE	0.99	1.13
AVERAGE PMPM 2016 COST	$794	$831
AVERAGE PMPM 2017 COST	$929	$978

Figures 22 and 23 show that, for the DxOPTml model in particular, Black people have fewer chronic conditions at each risk score percentile than white people. Given the results in prior sections using an age-matched sample, we believe it is likely that observed bias is influenced by differences in MARA's performance between the disabled population under age 65 and the age 65 and over population.

FIGURE 22: CHRONIC CONDITION COUNT BY RISK SCORE PERCENTILE, 5% SAMPLE, DxOPTml

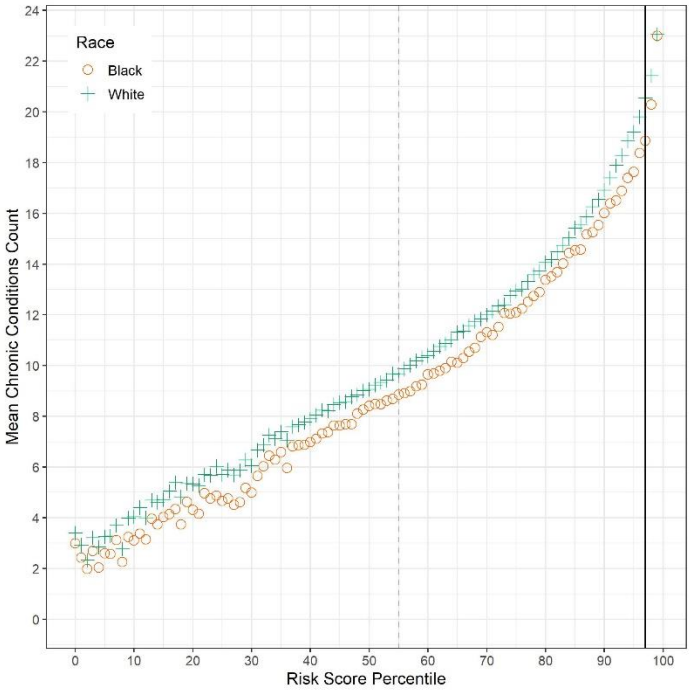
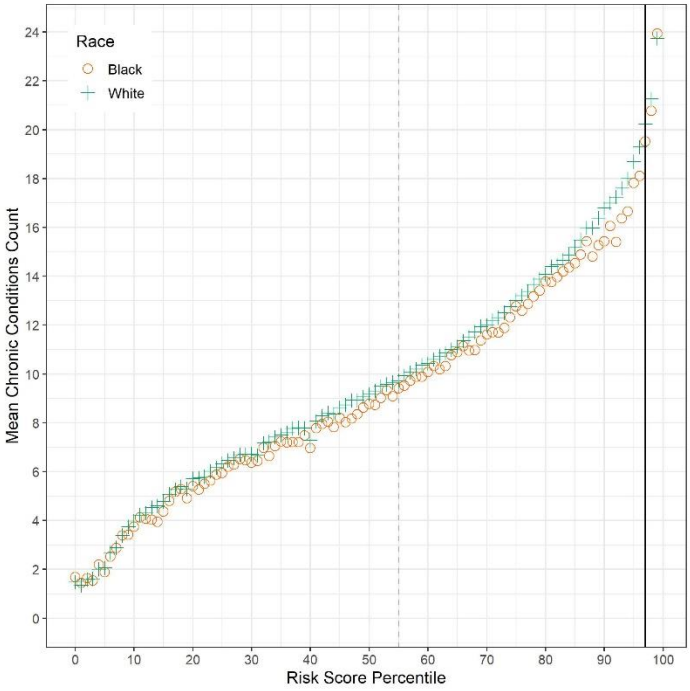


FIGURE 23: CHRONIC CONDITION COUNT BY RISK SCORE PERCENTILE, 5% SAMPLE, DxXPLN



The tables in Figures 24 and 25 similarly show that Black people in this sample generally have fewer chronic conditions than white people at the same level of risk. As described above, we do not think it is appropriate to draw any conclusions about the relationship between race and risk score performance given the material difference in ages between the samples.

FIGURE 24: CHRONIC CONDITION COUNTS, 5% SAMPLE, DxOPTml

	WHITE	BLACK	DIFFERENCE*
ALL MEMBERS	9.7	9.4	-3.3%
TOP 1%	23.0	23.0	-0.2%
TOP 3%	21.7	20.8	-3.9%
TOP 5%	20.8	19.8	-4.8%

* Calculated as (Black – White) / White.

FIGURE 25: CHRONIC CONDITION COUNTS, 5% SAMPLE, DxXPLN

	WHITE	BLACK	DIFFERENCE*
ALL MEMBERS	9.7	9.4	-3.3%
TOP 1%	23.7	23.9	0.8%
TOP 3%	21.7	21.5	-0.8%
TOP 5%	20.6	20.2	-2.0%

* Calculated as (Black – White) / White.

Milliman

Milliman is among the world's largest providers of actuarial and related products and services. The firm has consulting practices in life insurance and financial services, property & casualty insurance, healthcare, and employee benefits. Founded in 1947, Milliman is an independent firm with offices in major cities around the globe.

milliman.com

CONTACT

Erica Rode
erica.rode@milliman.com

Hans Leida
hans.leida@milliman.com