scientific reports

OPEN

# The association between BMI and healthcare burden, stratified by race and healthcare utilization among middle-aged patients in the US

Manal J. Alshakhs[1], Patricia J Goedecke[2], Lokesh K Chinthala[1], Nicole G. Weiskopf[3] & Charisse Madlock-Brown[1,4]✉

**This study assesses the relationship between BMI and healthcare burden, stratified by race and healthcare utilization, among middle-aged patients in the US. We used data from the Cerner HealthFacts database for 2016–2017 as our study period. We employed regression analysis to maximize the area under the curve (AUC) of the receiver operating characteristic (ROC) curve. This analysis assessed the relationship between BMI and healthcare burden using logistic regression and identified optimal BMI cutoffs stratified across different racial groups and levels of healthcare utilization by maximizing the AUC of the ROC curve. BMI was normalized between 0 and 1, and odds ratios were interpreted as the change in odds of a CCI score greater than zero associated with a full-range (min-to-max) increase in BMI. Results indicated that BMI was strongly associated with CCI across regular, low, and non-utilization cohorts. Specifically, in the regular healthcare utilizer cohort, a min-to-max increase in BMI was associated with a significantly higher likelihood of a CCI score greater than zero. The odds ratios for a min-to-max BMI change were notably high for the White and Asian/Pacific Islander groups (24.3 and 38.18, respectively), compared to 5.04 and 3.82 for the Black and Native American groups. The AUC analysis revealed the highest value for the Asian/Pacific Islander cohort (0.71) and the lowest for the Black cohort (0.63), with optimal BMI cutoffs identified as 34 for African Americans, 35 for American Indians/Alaska Natives, 32 for Whites, and 27 for Asians/Pacific Islanders. The findings underscore the necessity of stratifying patients by healthcare utilization, particularly as regular utilizers in the White and Asian/Pacific Islander populations had lower BMI cutoffs. This study advocates for a paradigm shift in obesity diagnosis, emphasizing the need for refined metrics and additional research on BMI's role in healthcare burden across diverse populations.**

**Keywords** Multimorbidity, Race stratification, Healthcare utilization, Healthcare burden

Body mass index (BMI) is a widely used measure in healthcare to assess an individual's weight status and determine their risk for developing health problems related to obesity. However, it is essential to note that BMI does not distinguish between muscle and fat, so it may overestimate body fat in athletes and underestimate body fat in older adults. Additionally, BMI does not consider individual differences in body shape and composition. The World Health Organization (WHO) has recommended rescaling the BMI for Asian individuals[1], based on studies performed in China that have produced new population-based BMI scales for obesity[2]. Indonesia, Turkey, Singapore, Japan, India, and other countries have done the same[3]. These BMI cutoffs vary depending on the data and resources available. Some measures used to set the new BMI cutoffs include adiposity, all-cause mortality, waist circumference, waist-to-hip ratio, and incidence rates of certain prevalent diseases[4–6]. A 2019 study also recommended new BMI cutoffs for Black, Hispanic, and White males and females, based on the risk of developing one of three metabolic diseases, namely dyslipidemia, diabetes mellitus, and hypertension.

[1]Health Outcomes and Policy Research Program, University of Tennessee Health Science Center, Memphis, TN, USA. [2]Bredesen Center, University of Tennessee, Knoxville, USA. [3]Department of Medical Informatics and Clinical Epidemiology, Oregon Health & Science University, Portland, OR, USA. [4]Community and Primary Care Division, College of Nursing, University of Iowa, 50 Newton Rd, Iowa City, IA 52242, USA. ✉email: cmadlock@uiowa.edu

These authors found differences in BMI cutoffs across race and gender for these three diseases[7]. The WHO recommends that population-based obesity BMI cutoffs be founded on morbidity and mortality[6]. Despite these developments, there remains a significant gap in knowledge for the United States. Previous studies have only identified race-specific BMI cutoffs in relation to particular diseases, which does not provide a comprehensive understanding of the general health risks associated with obesity. Additionally, population-specific BMI cutoffs have not been adequately researched or established for major racial groups within the U.S., particularly for Asian, Native American, and Black Americans.

Assessing the relationship between BMI and healthcare burden across race in the United States (U.S.) is crucial in determining a person's overall health status and informing the development of a tailored treatment plan. This relationship between BMI and healthcare utilization is generally positive, meaning that as BMI increases, so does healthcare utilization[8]. Higher BMIs are often associated with a greater risk of obesity-related health conditions such as type 2 diabetes mellitus, cardiovascular disease, and certain cancers[9]. These health conditions can lead to increased hospitalizations, medication use, and doctor visits, thereby straining the healthcare system. While individuals with lower BMIs are less likely to require as many healthcare resources, regular healthcare utilization results in better healthcare outcomes regardless of BMI status[10,11].

The Charlson Comorbidity Index (CCI) is a widely used tool for evaluating comorbidity in patients with chronic diseases and is often used to assess health outcomes[12]. This study aims to assess the relationship between BMI and CCI as stratified by race and healthcare utilization among middle-aged patients in the US. We hypothesize that BMI does not operate equally across race. This will help address health disparities and design target interventions to improve health outcomes.

## Materials and methods

### Research design

This cross-sectional study used data from the Cerner HealthFacts® data warehouse for 2016–2017 that did not contain any identifying information. The dataset comprises the encounter data from more than 490 million patient encounters with over 70 million patients who received treatment at hospitals and clinics from 792 non-affiliated healthcare systems across the U.S. from 2001 to 2017. The data includes information about the type of encounter, patient medical history, diagnoses, laboratory results, prescriptions, patient demographics, clinic type, and procedures performed. The Cerner HealthFacts® database did not capture waist circumference, which has strongly correlated with multimorbidity in previous research. However, this database did capture insurance and visit types. To ensure data quality, we included only healthcare centers that had recorded values for race, age, gender, and BMI for at least some patients. Of the 588 hospitals with data available during the study period, 474 met these criteria and were included in the analysis. We will use these two variables to investigate their association with the burden of multimorbidity across races.

For this study, the patient inclusion criteria were: (1) age 45–64 years, (2) BMI value present and ranging between 18.5 and 75, (3) assigned race category, and (4) assigned gender (i.e., male or female). The exclusion criteria were refined to omit specific patient groups where standard BMI analysis could be confounded by extenuating health conditions. These exclusions were: (1) patients with a cancer diagnosis, due to cancer and its treatments significantly affecting body weight either through loss or gain, thereby complicating the analysis of BMI and its direct relationship with other comorbidities; (2) pregnant patients, as pregnancy can temporarily alter weight and BMI in ways that do not reflect the patient's usual health status; and (3) patients with International Classification of Diseases-9th Version-Clinical Modification (ICD-9-CM) diagnostic codes. These exclusions were carefully chosen based on their potential impact on the variables under investigation, particularly BMI and its associations with other comorbidities. Cancer and its treatments can significantly affect body weight, either through loss or gain, which can obscure the analysis of BMI and its direct relationship with other comorbidities. Excluding these patients allows for a clearer examination of the BMI impacts uninfluenced by the extreme weight variations often seen in cancer patients. Pregnancy can temporarily alter weight and BMI in ways that do not reflect the patient's usual health status. Since BMI changes during pregnancy are typically not indicative of the individual's health outside of pregnancy, including pregnant patients could distort the analysis of BMI's effects on health outcomes. Using ICD-9-CM codes, due to their less specific nature than ICD-10-CM, could lead to inaccurate disease classifications affecting the study's precision in assigning comorbidity scores. CD-9 and ICD-10 differ not only in the number and structure of codes but also in the clinical details they capture. This disparity makes a direct or 1-to-1 mapping between the two systems challenging and often inaccurate. Given that our study relies heavily on precise coding to assign CCI values accurately, using two different coding systems could introduce a significant risk of misclassification and bias in our analysis. We selected a two-year study period to enhance our analytical sensitivity in capturing healthcare utilization patterns and to gather a more comprehensive dataset, ensuring robustness and depth in our findings.

We stratified the patients by race and divided them into three subgroups based on patterns of healthcare utilization, which we defined as a minimum of two outpatient visits within the two-year study period. Patients with no outpatient visits were classified as "non-utilizers," while those with only one such visit were classified as "low utilizers." It should be noted that all utilization subgroups may have had inpatient and/or emergency department visits. This decision was intentional and is grounded in our study's focus on regular preventive care. We chose to emphasize outpatient visits because they are a primary indicator of engagement in ongoing health maintenance, which can significantly reduce the burden of disease over time. Stratifying by utilization refers to grouping individuals or data based on their use of healthcare services. This can include the frequency of doctor visits, use of prescription medications, and hospitalization rates. By stratifying the data in this way, we can examine how different levels of healthcare utilization may be associated with various health outcomes. It is also crucial to stratify the data by race because some racial groups may have different levels of access to healthcare

that influence health-seeking behavior. By stratifying the data, we can examine how these differences in health-seeking behavior may alter associations with health outcomes for different racial groups.

### Ethical considerations

This research utilized a limited dataset that excluded the 16 identifiable variables. Per the National Institutes of Health policy, the University of Tennessee Health Science Center (UTHSC) Institutional Review Board (IRB) classified the research as exempt. This study was conducted in compliance with all other necessary research requirements.

Ethical considerations regarding data use agreements concerning specific American Indian/Alaska Native tribes are acknowledged. The study does not delve into specific tribes; therefore, we did not seek any data use agreements. We followed the recommendations for using American Indian/Alaska Native data by the Urban Indian Health Institute[13], which recommends not censoring these data as much as possible.

### Independent variables

Demographics, payer type, smoking, alcohol use, urbanism, and 2014–2015 healthcare utilization (i.e., patients who had an encounter during 2014–2015) were the primary independent variables. Demographic variables were race, age, gender, marital status, and body mass index (BMI), while racial categories were Black, Asian/Pacific Islander, White, and Native American. Given the limitations in sample size, we chose to include only these four racial categories due to their adequate representation within the Cerner HealthFacts dataset, despite the availability of additional racial categories. We categorized marital status as those patients currently with a partner (i.e., married or life partner), having an ex-partner (i.e., divorced, widowed, or legally separated), single, and unknown. We categorized payer types as those patients with public insurance, private insurance, self-pay, or unknown, based on the information available in the dataset. We limited BMI values to the range $18.5 \leq \text{BMI} < 75$, based on the distribution of this variable within our sample population. We defined 2014–2015 healthcare utilization as a patient having at least two outpatient visits during that period. In this study, the variables of age and BMI were standardized via min-max scaling to a uniform range of 0 to 1. This rescaling approach is critical in multivariate analyses as it normalizes the scales of variables with disparate measurement units and ranges, thereby ensuring that each variable contributes equitably to the regression model. By doing so, it mitigates the risk of introducing scale-induced bias into the model's estimations, thus enhancing the interpretability and numerical stability of the estimated parameters. Patients with unknown urban status were assigned to the urban category, since it had the highest prevalence in the dataset.

### Outcome variable

Our primary outcome variable was the healthcare burden measure, CCI, calculated using the Quan version of the Charlson Comorbidity Index[14] via the R comorbidity package from the International Classification of Diseases-10th Version-Clinical Modification (ICD-10-CM) diagnosis codes. The Quan version of the Charlson Comorbidity Index is an adaptation of the original CCI that has been enhanced for use with ICD-10 administrative data. This version improves the accuracy of comorbidity data capture by updating the coding algorithms to reflect changes in the classification and diagnosis of diseases, providing a more precise tool for assessing patient comorbidities in contemporary healthcare datasets. Thus, it is important to note that a patient might have morbidities or comorbidities that were not on this list. We chose CCI as our outcome variable since it is a reliable tool for evaluating health outcomes by quantifying the degree of illness in an individual patient[12] and provides a standardized measure of disease that allows patient comparisons over time[15]. We transformed CCI scores into a binary variable—0 for patients with a CCI of 0 and 1 for those with a CCI of 1 or higher. This modification was chosen to simplify the analysis and enhance the interpretability of data, particularly for identifying BMI cutoffs related to patients with any comorbidity burden. As we exclude patients with cancer, only the 16 other comorbidities in the Quan version of CCI are represented in our analysis.

### Statistical analysis

In this study, we applied logistic regression with a binomial family distribution to the training dataset (70% of the final population) to model the relationship between the predictor variables and the binary outcome variable, CCI[16]. The Kolmogorov-Smirnov (KS) test was performed to compare the distribution of the outcome variable among the different cohorts[17]. The results indicated no statistically significant deviation from normality for all the cohorts except for the White utilizer cohort, which the KS test showed to be statistically significant. We checked for multicollinearity, which refers to linear relationships between multiple variables[18] using a generalized variance inflation factor (GVIF) analysis suitable for a combination of numerical and categorical variables[19,20]. We used the car R package to remove variables with a high GVIF^(1/2Df ) score (i.e., alcohol use)[21] and repeated this process until no variable had a score above the conservative threshold of two[20].

We examined how BMI is associated with CCI score by comparing the area under the curve (AUC) of the receiver operating characteristic (ROC) curve for the adjusted model with and without inclusion of the BMI. We determined the best BMI cutoff for each group by analyzing the ROC curve and finding the point with the highest AUC within the range of 20 to 38 BMI for each group. We had 19 ROC curves for each cohort and considered only the BMI cutoff with the highest AUC.

## Results

In this study, the majority of patients in the combined cohorts were female and the average age was 55 years. Table 1 shows the breakdown of demographics by race. The percentages were determined in relation to the entire patient population. Table 2 shows the breakdown of the number of patients in different utilization cohorts by race.

| Variable | Black | Asian/Pacific Islander | White | Native American | Total |
|---|---|---|---|---|---|
| N (%) | 244,952 (19) | 23,151 (2) | 992,250 (78) | 11,344 (1) | 1,271,697 (100) |
| Gender n( %) | | | | | |
| Male | 104,743 (18) | 9727 (2) | 447,680 (79) | 5148 (1) | 567,298 (100) |
| Female | 140,209 (20) | 13,424 (2) | 544,570 (77) | 6196 (1) | 704,399 (100) |
| Age (Mean, SD) | 54 (5.6) | 54 (5.9) | 55 (5.7) | 54 (5.7) | 55 (5.7) |
| BMI (Mean, SD) | 32 (8.1) | 27 (5.6) | 31 (7.4) | 32 (7.7) | 31 (7.6) |
| Health insurance type n (%) | | | | | |
| Self-pay | 37,415 (25) | 2397 (2) | 112,079 (73) | 2337 (2) | 154,228 (100) |
| Public | 81,780 (27) | 3355 (1) | 212,905 (71) | 3801 (1) | 301,841 (100) |
| Private | 97,954 (14) | 12,870 (2) | 564,704 (83) | 4530 (1) | 680,058 (100) |
| Unknown | 27,803 (21) | 4529 (3) | 102,562 (76) | 676 (< 1) | 135,570 (100) |
| 2014–2015 healthcare utilization n(%) | | | | | |
| 2+ Outpatient visits in 2014–2015 | 83,104 (17) | 9137 (2) | 380,574 (80) | 3032 (1) | 475,847 (100) |
| < 2 Outpatient Visits in 2014–2015 | 161,848 (20) | 14,014 (2) | 611,676 (77) | 8312 (1) | 795,850 (100) |
| Smoking status n (%) | | | | | |
| Smoker | 38,351 (23) | 936 (1) | 125,968 (75) | 2061 (1) | 167,316 (100) |
| Non-smoker | 206,601 (19) | 22,215 (2) | 866,282 (78) | 9283 (1) | 1,104,381 (100) |
| Marital status | | | | | |
| Partnered | 74,113 (10) | 16,660 (2) | 589,766 (86) | 5082 (1) | 685,621 (100) |
| Ex-partner (divorced, widowed, or legally separated) | 41,685 (21) | 1815 (1) | 154,809 (77) | 2012 (1) | 200,321 (100) |
| Single | 124,186 (35) | 3634 (1) | 218,742 (62) | 3921 (1) | 350,483 (100) |
| Unknown martial status | 4968 (14) | 1042 (3) | 28,933 (82) | 329 (1) | 35,272 (100) |
| Healthcare setting urban/rural status | | | | | |
| Urban | 167,458 (18) | 11,547 (1) | 755,131 (80) | 7246 (1) | 94,1382 (100) |
| Rural | 77,494 (23) | 11,604 (3) | 237,119 (72) | 4098 (1) | 330,315 (100) |

**Table 1**. Patient demographics by race.

| Race/Utilization | Utilizers *n* (%) | Non-Utilizers *n* (%) | Low Utilizers *n* (%) | Total *n* (%) |
|---|---|---|---|---|
| Black | 111,161 (17) | 96,436 (25) | 37,355 (16) | 244,952 (19) |
| Asian/Pacific Islander | 13,665 (2) | 5152 (1) | 4334 (2) | 23,151 (2) |
| White | 526,115 (80) | 274,813 (72) | 191,322 (81) | 992,250 (78) |
| Native American | 4642 (1) | 4446 (1) | 2256 (1) | 11,344 (1) |
| **Total** | 655,583 (100) | 380,847 (100) | 235,267 (100) | 1,271,697 (100) |

**Table 2**. Number of patients in utilization cohorts by Race.

| Race/variable | Mean age (SD) | Mean BMI (SD) | Mean CCI (SD) |
|---|---|---|---|
| Black | 55 (5.5) | 33 (8.2) | 0.5 (0.8) |
| Asian/Pacific Islander | 54 (5.9) | 27 (5.4) | 0.3 (0.6) |
| White | 55(5.6) | 31 (7.6) | 0.3 (0.6) |
| Native American | 54 (5.7) | 32 (7.8) | 0.5 (0.7) |

**Table 3**. Mean and standard deviations for age, BMI, and CCI score by race for healthcare utilizer Cohorts.

The Cerner HealthFacts® database included data from 1,500,636 middle-aged patients for 2016–2017. The study flow diagram used to recruit the final patient population is shown in Appendix Fig. 1. After data cleaning procedures, excluding patients with ICD-9-CM diagnostic codes, patients with a diagnosis of cancer or pregnancy, and patients who received treatment in multiple urban locations, the remaining patient population comprised 1,271,697 individuals. The mean age, BMI, and CCI scores for each racial cohort are shown in Table 3, while the distribution of BMI across races is shown in Appendix Fig. 2.

In our logistic regression model, the Charlson Comorbidity Index (CCI) serves as the outcome variable, with predictors including BMI, age, gender, health insurance type, smoking status, marital status, healthcare utilization during 2014–2015, and urbanism. The reference categories for the categorical variables are as follows:

- **Gender**: Female.
- **Health Insurance Type**: Self-pay.
- **Marital Status**: Partnered.
- **Smoking Status**: Non-smoking.
- **Urbanism**: Rural.
- **Healthcare Utilization (2014–2015)**: Fewer than 2 outpatient visits.

Figures 1, 2 and 3 present the odds ratios for various health variables, stratified by race and healthcare utilization, displayed on a logarithmic scale. The logarithmic scale was employed to accommodate the significantly higher estimates associated with BMI compared to other variables, ensuring a clear and comprehensible visual representation across a broad range of values. Appendix Table 1 provides the non-logarithmically transformed estimates and their corresponding confidence intervals for each health variable. Based on the odds ratio, high BMI is significantly associated with the healthcare burden measure, CCI, in the Asian/Pacific Islander and White races across all utilization cohorts. In the healthcare utilizer cohort, BMI coefficient values almost doubled for those two races compared to the Black and the Native American races. BMI had the highest predictor of CCI across race and utilization. As the BMI variable was scaled between 0 and 1, the odds ratio represents the change in odds associated with moving from the lowest to the highest value of the BMI scale. This quantifies the relative risk increase or decrease for individuals at the high end of the BMI scale compared to those at the low end. The top three variables predicting CCI varied by healthcare utilization. For example, BMI, age, and smoking were the top three variables predicting CCI in the healthcare utilizer cohorts for all races. Yet, BMI, age, and public insurance were the top three variables predicting CCI in the healthcare non-utilizer cohorts for the Black and Native American races. At the same time, the rank remained the same for the remaining races. For the healthcare low utilizers, BMI and smoking were the top two variables positively predicting CCI in the Black race. The third

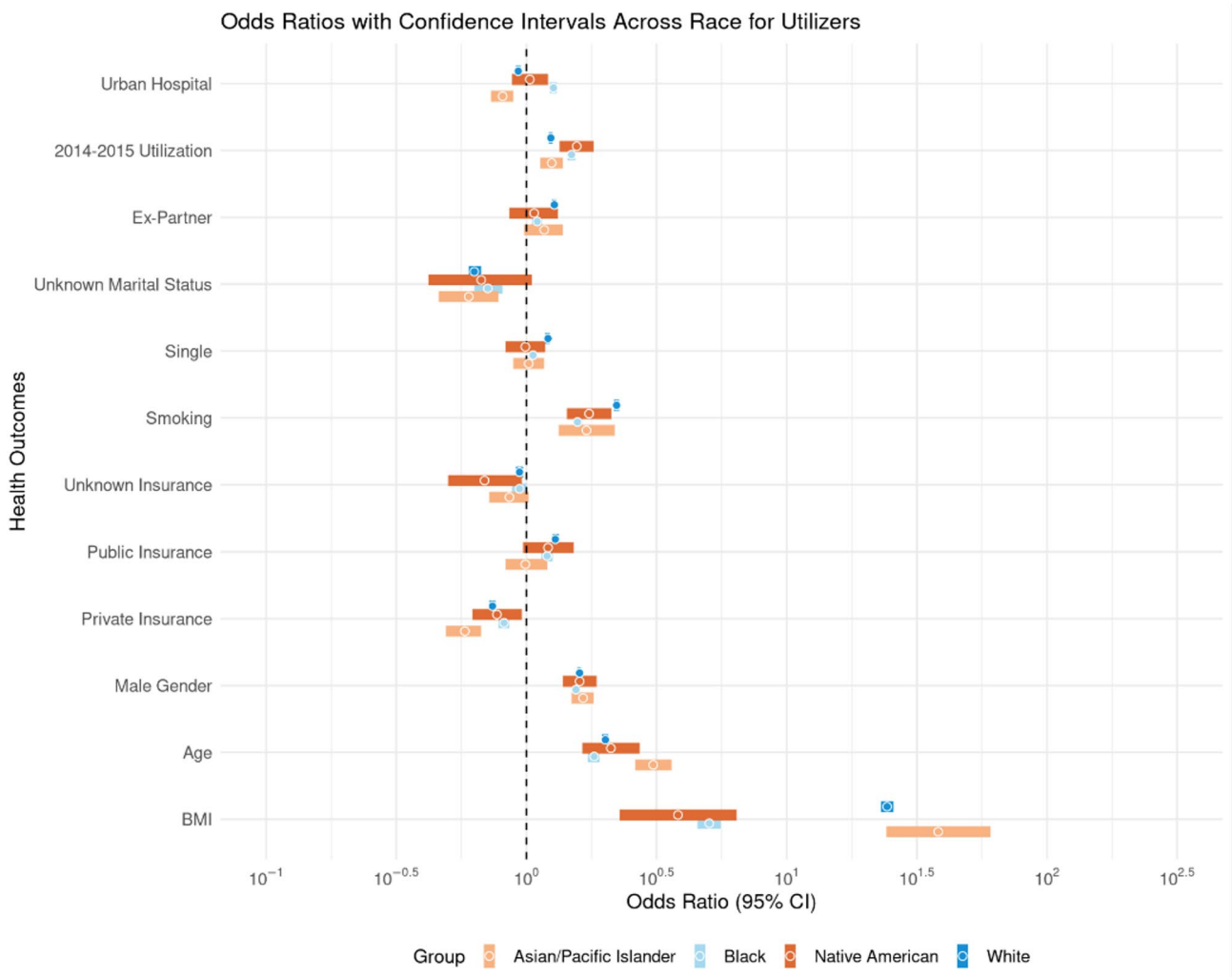


**Fig. 1**. Odds Ratios with Confidence Intervals Across Racial Groups for Utilizers. Estimates that are not statistically significant cross the reference line at $10^0$ on the y-axis. The "Ex-partner" variable represents patients who are divorced, widowed, or legally separated.

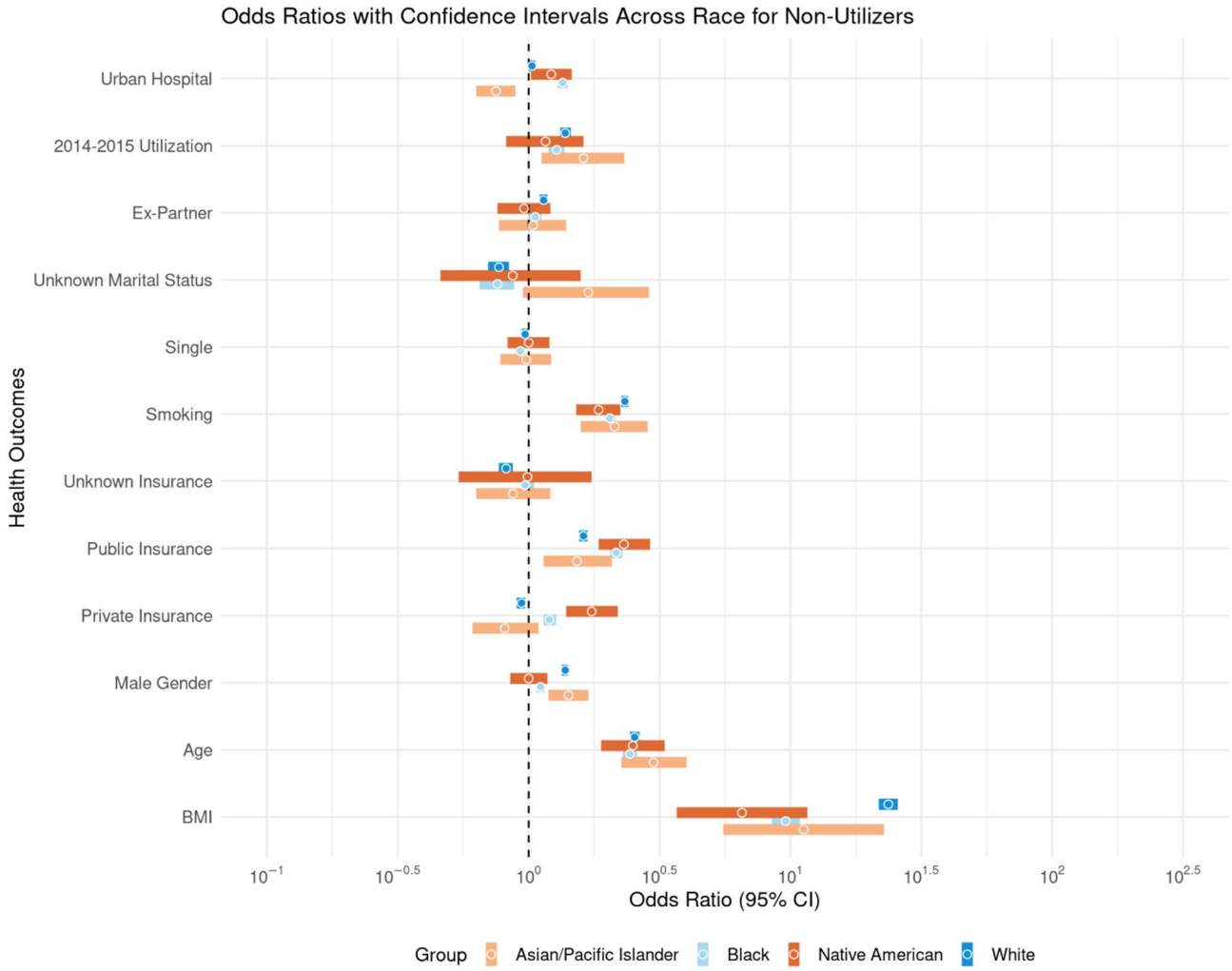


**Fig. 2**. Odds Ratios with Confidence Intervals Across Race for Non-Utilizers. Estimates that are not statistically significant cross the reference line at $10^0$ on the y-axis. The "Ex-partner" variable represents patients who are divorced, widowed, or legally separated.

was unknown marital status, which was negatively associated with CCI prediction. Age, smoking, and BMI remained the top three predictors for the remaining races in the low-utilizer cohorts.

We assessed the association of BMI with CCI score by displaying the AUC of the ROC curve for the adjusted model with and without BMI in the health utilizer cohort alongside the highest AUC of the adjusted model with BMI in the 20–38 range, as shown in Fig. 4. The highest AUC was achieved for the Asian/Pacific Islander cohort (0.71) and the lowest for the Black cohort among the healthcare utilizer cohort (0.63). The cutoff for the highest AUC in the healthcare utilizer cohorts was 34 for the Black cohort, 27 for the Asian/Pacific Islander cohort, 32 for the White cohort, and 35 for the Native American cohort. The assessment of the remaining utilization cohorts is shown in Appendix Figs. 3 and 4. In the low utilization cohort, the highest AUC was achieved for the White racial group, followed by that for the Asian/Pacific Islander racial group, with the CCI for the Black racial group being predicted least well by the model. In the non-utilization cohort, AUC values range between 0.65 and 0.69, with the Native American cohort having the highest value and Asians/Pacific Islanders having the lowest. BMI cutoffs vary differently based on utilization. For the Asian/Pacific Islander Cohort, BMI cut-offs were higher in low and non-utilizer cohorts and lower for African Americans in the low-utilizer cohort than the regular healthcare usage cohort.

## Discussion

This study assessed the relationship between BMI and the Charlson Comorbidity Index (CCI) across different racial groups. BMI was significantly associated with CCI, regardless of healthcare utilization and race. The AUC values suggest variations in the model's ability to predict CCI, with an optimal cutoff for African American and Native American populations being higher than for White populations. This data can help researchers and healthcare providers understand and address health disparities[22–25]. Additionally, the variables associated with CCI varied depending on the degree of healthcare utilization being analyzed. For the White and Asian/Pacific Islander populations, low utilizers of regular care demonstrated higher odds ratios for BMI. As racial

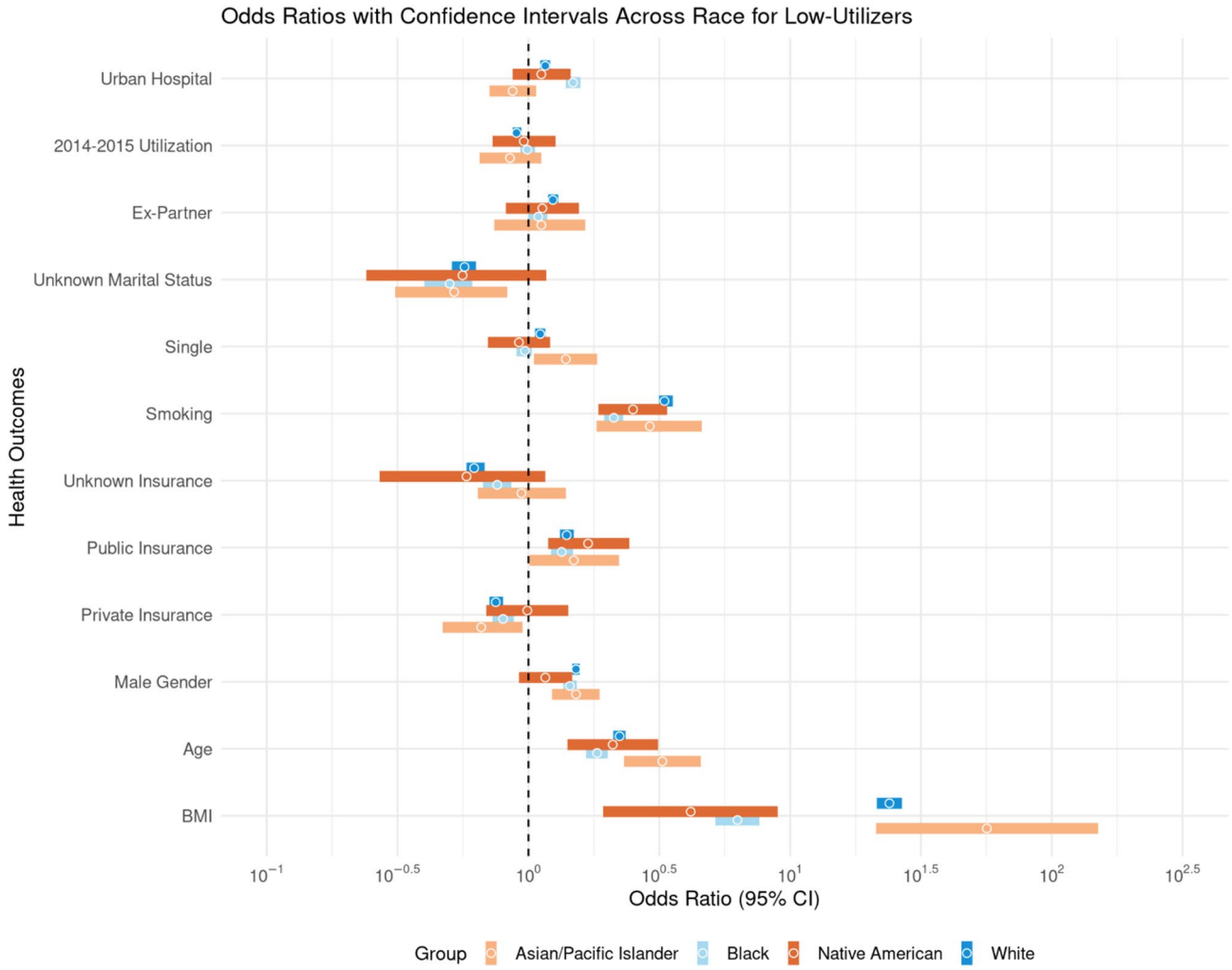


**Fig. 3**. Odds Ratios with Confidence Intervals Across Race for Low-Utilizers. Estimates that are not statistically significant cross the reference line at $10^0$ on the y-axis. The "Ex-partner" variable represents patients who are divorced, widowed, or legally separated.

minority populations tend to have greater utilization of emergency services[26], it is important to factor utilization into this kind of research. Users of "real world" EHR data should be aware that utilization can create bias in obesity research, and stratification is necessary. This research supports the growing emphasis from the American Medical Association on understanding the limitations of BMI[27]. Our findings across race groups reinforce the study's central argument that relying solely on BMI without considering racial nuances can lead to suboptimal predictions of health outcomes.

Our study aligns with the growing movement highlighted in the work by Prillaman to move beyond an overreliance on BMI in obesity diagnosis and treatment[28]. The emphasis on racial differences adds a critical layer to this discussion, acknowledging that a one-size-fits-all approach is inadequate. In the context of the global demand for anti-obesity medications, our study raises concerns about the potential for BMI to continue as a predominant diagnostic tool, leading to the prescription of medications without thorough consideration of individual health contexts. This aligns with the Nature paper's cautionary note about the impact of increasing reliance on BMI in the era of emerging weight-loss drugs. Studies consistently show that people from different ethnic and racial backgrounds may have different body compositions for a given BMI. For example, individuals of Asian descent may have a higher percentage of body fat at lower BMI values compared to individuals of European descent[2]. Some research suggests that health risks associated with obesity-related conditions may manifest at lower BMI levels in certain non-white populations compared to white populations[7]. Some countries and regions have started developing or adopting population-specific guidelines for assessing obesity and related health risks. These guidelines consider the unique characteristics of their populations, including genetic, cultural, and lifestyle factors. Our work adds to this mounting evidence by identifying the association between BMI and a common comorbidity burden score to identify cutoffs for BMI across racial groups in the United States. By factoring in utilization, we were able to control for a key characteristic that may bias results in other studies.

The study's findings prompt reflection on the broader implications for clinical practice. While maximizing the AUC to identify BMI cutoffs represents a statistical optimization, its clinical relevance requires further

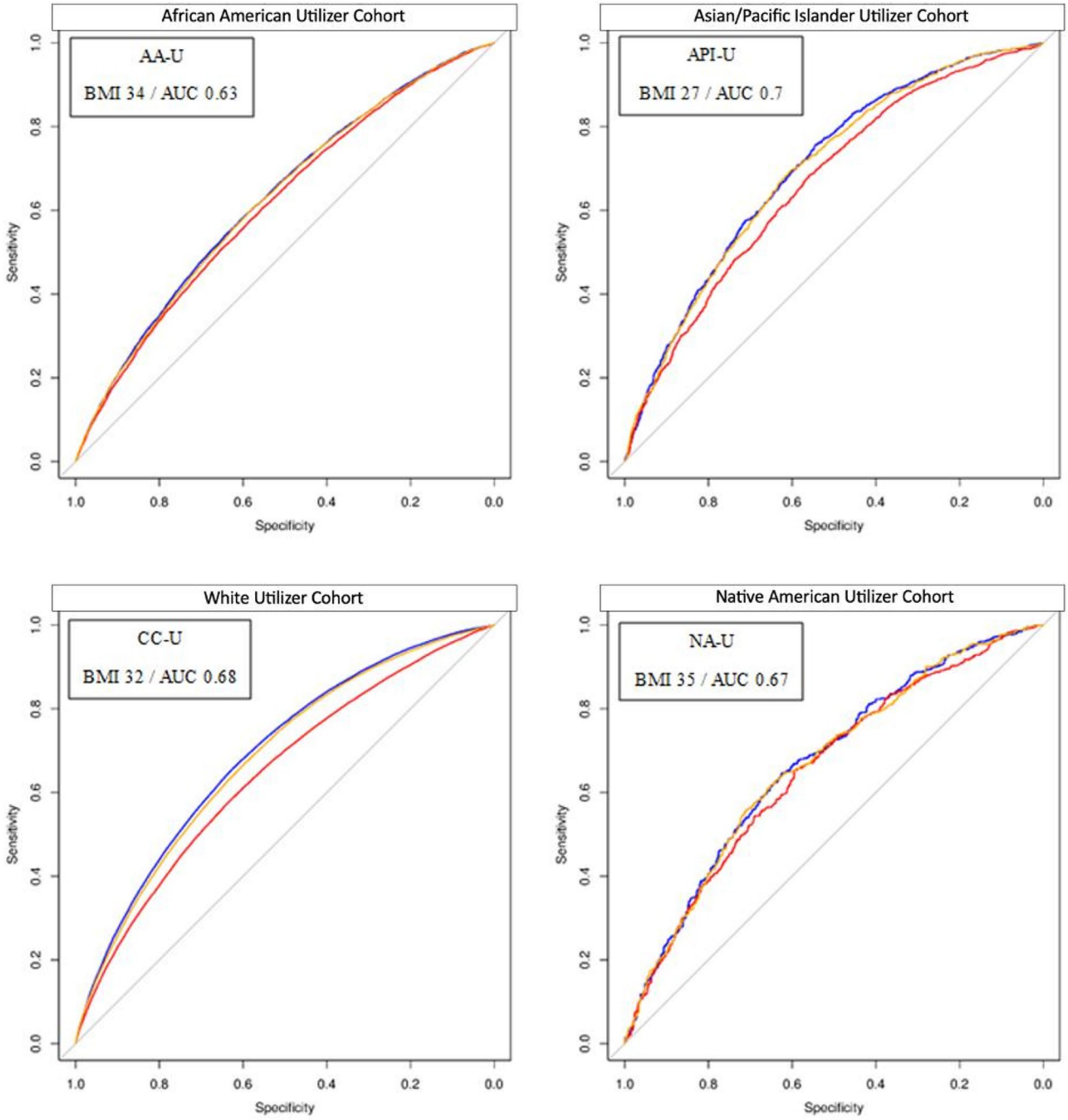


**Fig. 4**. Comparing Area Under the Curve (AUC) for different models to evaluate the association between BMI and CCI for healthcare utilizer cohorts. Red curves represent the unadjusted model without BMI, the blue curves represent the adjusted model with BMI, and the orange curves represent the highest AUC achieved and the corresponding BMI cutoff for that AUC value.

investigation. Future research should focus on translating these statistical findings into meaningful clinical interventions and guidelines.

## Limitations

A cross-sectional design may not fully capture the relationship between BMI, healthcare burden, and healthcare utilization complexities. It may not control for other important factors influencing healthcare utilization, such as genetics, lifestyle, and overall health. Additionally, the inherent limitations of a cross-sectional study restrict our ability to support causal assumptions, as it captures data at a single point in time rather than across different time periods. This prevents us from establishing temporal sequences and causality between the observed variables. Many patients were excluded from the analysis due to missing BMI, with a rate of 44% missingness. Ad hoc

analysis, as shown in Appendix Table 2 was conducted to compare patients with and without recorded BMI values across several key variables: age, race, sex, visit count, and insurance status. Our findings confirm that there are no significant differences in average age (45 vs. 45) or the percentage of white individuals (70% vs. 71%) between our study cohort and the cohort without BMI values, respectively. Furthermore, our study cohort shows a more balanced gender distribution (55% female and 45% male vs. 68% female and 39% male), closer to the real distribution in the United States. However, our study cohort exhibits a higher percentage of Black individuals (18% vs. 12%) and a lower rate of missing race data (2.9% vs. 8.2%) This observed difference in missingness might obscure the true proportion of Black individuals and is likely associated with utilization. Notably, the average visit rate in our study cohort is more than twice that of the cohort without BMI values (19 vs. 7) during the study period. Higher utilization is likely associated with the presence of a BMI value as well as the more visits a patients has the more opportunities to have this value collected. Additionally, fewer patients in the study cohort lack a recorded payer type, and a greater percentage have multiple payer types, which likely corresponds to their higher average visit count. The impact of these differences driven by utilization are controlled for by stratifying by utilization in our study[29]. We also limited our study to middle-aged patients, meaning our conclusions cannot be generalized to the whole population. Cerner HealthFacts had a limited number of variables that we could control for. Future work that includes more variables such as education, physical activity, stress levels, and diet, and includes more younger and older age groups over a longer period of time could provide a more comprehensive approach for understanding the BMI-CCI relationship.

## Conclusions

To our knowledge, this is the first study to assess the relationship between BMI and healthcare burden as stratified by race and healthcare utilization. This research demonstrated that the relationship between BMI and CCI varied across race within the same healthcare utilization cohorts. Some of this variation could be driven by access to healthcare resources. Most prior research regarding multimorbidity focused on a specific point in time and a few multimorbidities at the same time. Understanding how multimorbidity accumulates over time across populations has not been addressed. More work needs to be done to understand how multimorbidity, BMI, and healthcare burdens are associated across race.

## Data availability

The data that support the findings of this study are available from Cerner Health Facts, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are, however, available from the corresponding author upon reasonable request and with the permission of [Cerner Health Facts].

Received: 25 October 2024; Accepted: 17 June 2025
Published online: 01 July 2025

## References

1. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *The Lancet*. **363** (9403), 157–163 (2004).
2. He, W. et al. Lower BMI cutoffs to define overweight and obesity in china: Chinese BMI cutoffs. *Obesity* **23** (3), 684–691. 10/f3pg2s (2015).
3. Deurenberg-Yap, M. & Deurenberg, P. Is a Re-evaluation of who body mass index Cut-off values needed?? The case of Asians in Singapore. *Nutr. Rev.* **61** (suppl_5), S80–S87. 10/c4gmb5 (2003).
4. Misra, A. Ethnic-Specific criteria for classification of body mass index: A perspective for Asian Indians and American diabetes association position statement. *Diabetes Technol. Ther.* **17** (9), 667–671. https://doi.org/10.1089/dia.2015.0007 (2015).
5. Wildman, R. P., Gu, D., Reynolds, K., Duan, X. & He, J. Appropriate body mass index and waist circumference cutoffs for categorization of overweight and central adiposity among Chinese adults. *Am. J. Clin. Nutr.* **80** (5), 1129–1136. https://doi.org/10.1093/ajcn/80.5.1129 (2004).
6. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *Lancet* **363** (9403), 157–163 (2004).
7. Stanford, F. C., Lee, M. & Hur, C. Race ethnicity, sex, and obesity: Is it time to personalize the scale? *Mayo Clin. Proc.*. **94** (2), 362–363 (2019). https://doi.org/10.1016/j.mayocp.2018.10.014
8. Ogden, C. L., Carroll, M. D., Kit, B. K. & Flegal, K. M. Prevalence of obesity and trends in body mass index among US children and adolescents, 1999–2010. *JAMA* **307** (5), 483–490. https://doi.org/10.1001/jama.2012.40 (2012).
9. Obesity and overweight. https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight. Accessed 9 Feb 2023.
10. Shiroma, E. J. & Lee, I. M. Physical activity and cardiovascular health: lessons learned from epidemiological studies across age, gender, and race/ethnicity. *Circulation* **122** (7), 743–752. https://doi.org/10.1161/CIRCULATIONAHA.109.914721 (2010).
11. Hunger, J. M., Smith, J. P. & Tomiyama, A. J. An Evidence-Based rationale for adopting Weight-Inclusive health policy. *Social Issues Policy Rev.* **14** (1), 73–107. https://doi.org/10.1111/sipr.12062 (2020).
12. Charlson, M. E., Carrozzino, D., Guidi, J. & Patierno, C. Charlson comorbidity index: A critical review of clinimetric properties. *PPS* **91** (1), 8–35. https://doi.org/10.1159/000521288 (2022).
13. Urban Indian Health Institute.Best Practices for American Indian and Alaska Native Data Collection. https://aipi.asu.edu/sites/default/files/best-practices-for-american-indian-and-alaska-native-data-collection.pdf (2023).
14. Quan, H. et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med. Care*. **43** (11), 1130–1139 (2005).
15. Jassal, S. V., Schaubel, D. E. & Fenton, S. S. A. Baseline comorbidity in kidney transplant recipients: A comparison of comorbidity indices. *Am. J. Kidney Dis.* **46** (1), 136–142. https://doi.org/10.1053/j.ajkd.2005.03.006 (2005).
16. Logistic Regression. Accessed August 21. (2021). https://www.sheffield.ac.uk/polopoly_fs/1.233565!/file/logistic_regression_using_SPSS_level1_MASH.pdf
17. 1.3.5.16. Kolmogorov-Smirnov Goodness-of-Fit Test. https://www.itl.nist.gov/div898/handbook/eda/section3/eda35g.htm. Accessed 2 Feb 2023.
18. Alin, A. Multicollinearity. *WIRE Comput. Stat.* **2** (3), 370–374. https://doi.org/10.1002/wics.84 (2010).

19. Allison, P. When can you safely ignore multicollinearity? Statistical horizons. Accessed 14 Nov 2022. https://statisticalhorizons.com/multicollinearity/
20. Fox, J. & Monette, G. Generalized collinearity diagnostics. *J. Am. Stat. Assoc.* **87** (417), 178–183. https://doi.org/10.2307/2290467 (1992).
21. Fox, J. & Weisberg, S. *An R Companion to Applied Regression* (SAGE, 2018).
22. National Academies of Sciences. *Factors That Affect Health-Care Utilization*. (National Academies Press (US), 2018). https://www.ncbi.nlm.nih.gov/books/NBK500097/. Accessed 21 Jan 2023.
23. Dashputre, A. A., Surbhi, S., Podila, P. S. B., Shuvo, S. A. & Bailey, J. E. Can primary care access reduce health care utilization for patients with obesity-associated chronic conditions in medically underserved areas? *J Eval Clin. Pract.* **20**. https://doi.org/10.1111/jep.13360 (2020).
24. McCormack, L. A., Jones, S. G. & Coulter, S. L. Demographic factors influencing nonurgent emergency department utilization among a medicaid population. *Health Care Manag Sci.* **20** (3), 395–402. https://doi.org/10.1007/s10729-016-9360-8 (2017).
25. Glynn, L. G. et al. The prevalence of Multimorbidity in primary care and its effect on health care utilization and cost. *Fam. Pract.* **28** (5), 516–523. https://doi.org/10.1093/fampra/cmr013 (2011).
26. Parast, L. et al. Racial/Ethnic differences in emergency department utilization and experience. *J. Gen. Intern. Med.* **37** (1), 49–56 (2022).
27. American Medical Association. AMA: Use of BMI alone is an imperfect clinical measure. https://www.amaassn.org/delivering-care/public-health/ama-use-bmi-alone-imperfect-clinical-measure (2023).
28. Prillaman, M. Why BMI is flawed — and how to redefine obesity. *Nature* **11** (7982), 232–233 (2023).
29. Allison, P. D. Missing data. In *Handbook of Survey Research*, 2nd ed. (eds Marsden, P. V. & Wright, J. D.) (Emerald Group Publishing Limited, 2010).



## Acknowledgements
We want to thank the UTHSC Center for Biomedical Informatics for the use of the Cerner HealthFacts dataset. We would also like to thank Elizabeth Elias for contributions to the final formatting of this paper.


## Author contributions
M.A.: conception, analysis, methods, data collection, writing. P.G: design, review. L.C.: data collection, review. N.W.: methods, design. C.M.B.: conception, design, methods, supervision, critical review.


## Funding
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.


## Declarations

## Competing interests
One of our co-authors does have potential competing interests. Nicole G. Weiskopf has provided consulting for Merck & Co., Inc. Merck had no role in this work. All other co-authors have no competing interests.

## Ethical approval
Cerner Health Facts has been certified by the UTHSC IRB to meet the conditions of "not human subjects research" and does not require IRB review or any interaction with the IRB. All methods were carried out in accordance with relevant guidelines and regulations. Research study and use of Cerner Health Facts data has been approved by UTHSC.

## Consent to participate
Informed consent was not required per UTHSC IRB guidelines.

## Additional information
**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-07779-9.

**Correspondence** and requests for materials should be addressed to C.M.-B.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025