

Research Article

# Health Informatics Analysis of Respiratory Disease Disparities Among Diverse Patient Populations

**Philip de Melo\*** , **Mildred Jackson, Aleyasia James, Keanna Lawrence, Shakira McElveen, Heaven Mitchell, Angel Tolbert-Abdullah, Ashley Wallace, Jasmine Waller**

Department of Nursing and Allied Health, Norfolk State University, Norfolk, United States of America

## Abstract

This study examines racial variation in length of stay (LOS) among patients hospitalized for respiratory illness in Texas to determine whether observed differences represent true health disparities or reflect underlying demographic and structural factors. Using population health principles, the analysis compares both total LOS and average LOS across racial groups, evaluates the impact of missing race data, and assesses age distribution as a potential confounding variable. Findings indicate that total LOS disproportionately reflects population size, with White patients demonstrating the highest aggregate LOS largely due to greater representation in older age groups. However, when LOS is measured at the individual level using average LOS, differences between racial groups become less pronounced. The analysis also identifies the “Null” race category as a source of missing demographic data that may introduce bias and affect the validity of disparity assessments. Additionally, age distribution patterns suggest that age significantly influences LOS outcomes and may confound interpretations of racial disparities. These findings demonstrate that raw differences in healthcare outcomes should not automatically be interpreted as evidence of inequity without adjustment for demographic and contextual variables. The study highlights the importance of using standardized, age-adjusted, and population-sensitive measures in population health analytics to ensure accurate interpretation of disparities and support equitable healthcare decision-making in Texas.

## Keywords

Length of Stay, Health Disparity, Confounders, Respiratory Diseases

## 1. Introduction

Health disparities remain a significant public health concern in the United States, particularly among populations affected by differences in access to healthcare, socioeconomic conditions, and demographic characteristics. In Texas, respiratory illnesses contribute substantially to hospitalization rates and healthcare utilization, making the examination of patient outcomes across racial groups an important area of population

health research. One commonly used indicator of healthcare burden is hospital length of stay (LOS), which reflects the duration of inpatient care and may signal differences in disease severity, access to treatment, quality of care, or underlying social determinants of health. However, variation in LOS does not automatically indicate the presence of a health disparity, as demographic and structural factors may also influence observed outcomes.

\*Correspondence: Philip de Melo ([pdmelo@nsu.edu](mailto:pdmelo@nsu.edu))

**Received:** 14 May 2026; **Accepted:** 25 May 2026; **Published:** 30 May 2026



Copyright: © The Author(s), 2026. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

Population health frameworks emphasize the importance of evaluating health outcomes within the broader context of social, economic, and environmental influences. According to Nash et al. [1] disparities should be interpreted carefully and supported by analyses that account for confounding variables rather than relying solely on descriptive comparisons. Measures such as total LOS may be heavily influenced by population size, while average LOS may provide a more accurate representation of individual patient experiences. In addition, incomplete demographic information and unequal age distribution across racial groups may further complicate the interpretation of disparities.

This study analyzes respiratory illness hospitalization data in Texas to explore differences in LOS across racial groups and determine whether these differences represent true health disparities. The analysis compares total and average LOS, examines the impact of missing race data, and evaluates age distribution as a potential confounding variable. By applying population health principles and considering demographic context, this study aims to provide a more accurate interpretation of healthcare variation and contribute to evidence-based approaches for addressing health inequities in Texas.

Health disparities are defined as preventable differences in health outcomes that are closely linked with social, economic, and environmental disadvantage. Public health researchers have consistently demonstrated that disparities are influenced by factors such as race, income, education, healthcare access, and living conditions. Population health focuses on understanding patterns of disease and healthcare outcomes across groups while examining the broader determinants that contribute to inequities. This framework emphasizes that statistical differences alone are insufficient to establish disparities unless they are connected to structural disadvantages and unequal opportunities for health.

Length of stay (LOS) is widely used in healthcare research as an indicator of disease severity, healthcare utilization, and quality of care. Previous studies have shown that LOS may vary across racial and ethnic populations due to differences in chronic disease burden, delayed access to care, socioeconomic status, and healthcare resources. However, researchers caution that aggregate measures such as total LOS may produce misleading conclusions because they are strongly influenced by population size. Public health literature recommends the use of standardized or per-capita measures, such as average LOS, to allow meaningful comparisons between demographic groups.

Research on healthcare disparities also highlights the importance of demographic variables such as age when analyzing patient outcomes. Older adults generally experience longer hospital stays due to higher rates of chronic illness, multiple comorbidities, and increased disease severity. Unequal age distribution across racial groups may therefore confound analyses of LOS if not properly controlled. Nash et al. (2019) explain that confounding variables can distort the rela-

tionship between demographic characteristics and health outcomes, leading to inaccurate conclusions about disparities. As a result, public health researchers often use age-adjusted rates, stratified analyses, or multivariate statistical methods to isolate the effects of race and other social determinants on healthcare outcomes.

Another significant issue identified in the literature is the impact of missing demographic data on disparity analysis. Incomplete race and ethnicity reporting can bias results and reduce the accuracy of population health assessments. Studies suggest that missing data may not occur randomly and may instead reflect systemic issues such as inconsistent reporting practices, mistrust of healthcare systems, or barriers to data collection.

Recent literature on respiratory illness outcomes further supports the need for careful interpretation of racial variation in healthcare utilization. Studies conducted during periods of increased respiratory disease burden, including influenza and COVID-19 outbreaks, revealed substantial differences in hospitalization patterns among racial and ethnic groups. However, these differences were often associated with broader social determinants of health, including housing density, occupational exposure, insurance coverage, and access to preventive care. Researchers emphasize that identifying true disparities requires moving beyond descriptive statistics to examine the structural and contextual factors that influence outcomes.

Overall, the literature demonstrates that evaluating health disparities requires a multidimensional approach that incorporates demographic adjustment, accurate data collection, and consideration of social determinants of health. Existing research supports the use of average LOS and adjusted analyses rather than relying solely on aggregate measures. The literature also underscores the importance of addressing missing data and confounding variables, particularly age, to ensure accurate interpretation of racial differences in healthcare outcomes. These findings provide the foundation for the present study examining respiratory illness hospitalization disparities in Texas.

Health disparities remain a major concern in public health and sports medicine, particularly among populations experiencing unequal access to healthcare, recreation, and wellness opportunities. Recent literature has increasingly emphasized the need to identify structural inequities that contribute to differences in injury outcomes, healthcare utilization, mortality, and participation in physical activity across racial, ethnic, and disability groups. Kliethermes et al. [2] conducted a comprehensive scoping review of North American-based health disparity research in sports medicine and found that most studies focused on racial and socioeconomic inequities, with growing attention to gender identity, disability, and access to care. The authors highlighted substantial methodological inconsistencies across studies and emphasized the need for standardized frameworks to improve health disparity research in sports medicine.

Similarly, Baker et al. [3] argued that disparities in sport and

recreation injuries are often overlooked despite their significant impact on vulnerable populations. Their study stressed that social determinants of health, including income, race, education, and neighborhood environment, directly influence injury risk and recovery outcomes. Tucker et al. [4] further expanded this discussion during the COVID-19 pandemic by demonstrating how healthcare inequities disproportionately affected athletes and performers with limited healthcare access, thereby exposing weaknesses in traditional sports medicine systems. Tuakli-Wosornu et al. [5] reinforced the concept of sports equity by advocating for inclusive policies and equitable healthcare delivery models that prioritize underserved populations in athletic and recreational settings.

Research examining racial and ethnic inequalities in broader health outcomes has also contributed significantly to understanding health disparities. Puka et al. [6] investigated whether lifestyle factors could explain racial and ethnic differences in all-cause mortality among U.S. adults. Their findings indicated that although behaviors such as smoking, diet, and physical activity contribute to mortality differences, these factors alone do not fully explain disparities, suggesting that structural and socioeconomic determinants remain major contributors to unequal health outcomes. This aligns with public health theories that disparities emerge not solely from individual behaviors but from systemic inequities embedded within healthcare and social institutions.

Several studies have focused on inclusion and participation in physical activity among marginalized populations. Bopp et al. (2017) explored perceived welcomeness in sports participation among racial and ethnic groups and found that minority participants frequently reported lower levels of inclusion and belonging in recreational sport environments. Brown et al. [7] examined motivations among midlife Black female triathletes and identified social support, empowerment, and representation as important influences on sustained physical activity participation. Calzo et al. [8] similarly identified disparities in physical activity participation among sexual minority youth, noting that gender nonconformity and reduced athletic self-esteem contributed to lower activity engagement among these populations.

Disability-related disparities also represent an important area of health equity research. Fagher et al. [9] emphasized the importance of para sport in promoting health equity among individuals with disabilities and argued that adaptive sports provide opportunities for physical, psychological, and social well-being. Baria et al. [10] investigated experiences among wrestlers with limb deficiencies and highlighted the importance of inclusive athletic opportunities and specialized healthcare support. Likewise, Barfield and Malone [11] examined barriers and benefits associated with power wheelchair soccer participation and found that adaptive sports improved quality of life despite persistent environmental and accessibility barriers. Braza et al. [12] further demonstrated the value of

academic-community partnerships in promoting awareness of adaptive sports opportunities to underserved populations.

Health disparities extend beyond athletics into broader patterns of social participation and community engagement. Budavari et al. [13] studied adolescents with autism spectrum disorder and found significant disparities in recreational and community participation compared to neurotypical peers. These findings suggest that social exclusion and limited access to inclusive activities contribute to broader inequities in health and quality of life. Hartill and Rulofs [14] also highlighted disparities in exposure to interpersonal violence in sports settings across European countries, emphasizing the need for protective policies and equitable safeguarding practices for children participating in athletics.

Methodologically, several foundational studies guide the conduct of health disparity reviews and population health research. Arksey and O'Malley [15] introduced the scoping review framework widely used in health sciences to map evidence across emerging areas of research. Peters et al. [16] later updated methodological guidance for scoping reviews, emphasizing transparency, reproducibility, and systematic reporting procedures. Tricco et al. [17] developed the PRISMA-ScR checklist to standardize reporting practices for scoping reviews and improve methodological rigor. Additionally, Bramer et al. [18] addressed database de-duplication methods for systematic reviews, improving the reliability and efficiency of evidence synthesis in health research.

Emerging technologies and advanced analytics have also been applied to healthcare disparity research. De Melo et al. [19] demonstrated the use of Kaplan-Meier analytics for accurate survival prediction in cancer research, illustrating how predictive modeling can support improved healthcare decision-making. De Melo and St. Rose [20] further explored the use of generative artificial intelligence in diabetes classification, highlighting the potential of machine learning technologies to enhance disease prediction and population health management. These studies suggest that advanced analytics may play an increasingly important role in identifying at-risk populations and addressing healthcare inequities through precision-based interventions.

Collectively, the literature demonstrates that health disparities are multidimensional and influenced by race, disability status, socioeconomic factors, healthcare access, and social inclusion. Existing studies consistently emphasize the importance of equitable healthcare systems, inclusive recreational opportunities, and evidence-based public health interventions. Furthermore, methodological advancements in scoping reviews, data analytics, and predictive modeling provide valuable tools for improving future health disparity research. These findings support the importance of continued investigation into healthcare inequities and reinforce the need for targeted strategies that promote equitable health outcomes across diverse populations.

## 2. Data Description

### 2.1. Overview of the Dataset

The Texas respiratory illness dataset was developed to support population health analysis of hospitalization outcomes across demographic groups within the state of Texas. The dataset contains patient-level hospitalization information associated with respiratory illnesses and was used to evaluate healthcare utilization patterns, length of stay (LOS), and demographic variation among hospitalized populations. The data provide an opportunity to explore potential health disparities and identify factors influencing patient outcomes across racial and age groups.

The dataset includes several demographic and clinical variables relevant to respiratory illness hospitalizations. Key variables examined in this study include race, age, and hospital length of stay. Race was used as the primary grouping variable to evaluate differences across populations, while age served as an important demographic factor for assessing potential confounding effects. Length of stay was used as the primary outcome measure because it reflects healthcare utilization, disease severity, and resource consumption within hospital systems.

### 2.2. Data Structure and Variables

The dataset is organized in tabular format, with each row representing an individual hospitalization encounter. Variables included in the dataset capture demographic characteristics and healthcare outcome measures associated with respiratory illness admissions in Texas hospitals. Important variables used in the analysis include:

- 1) Race/Ethnicity
- 2) Age
- 3) Length of Stay (LOS)
- 4) Hospitalization Counts
- 5) Patient Classification Categories

The race variable categorizes patients into multiple demographic groups, including White, Black, Asian or Pacific Islander, American Indian, Other, and a “Null” category representing missing or unreported race information. This variable was central to the disparity analysis because racial differences in healthcare outcomes are frequently associated with social determinants of health, healthcare access, and structural inequities.

The age variable was grouped into categorical ranges to examine population distribution patterns across racial groups. Age is an important factor because older populations are generally associated with increased disease burden, chronic illness prevalence, and longer hospital stays. Therefore, age distribution was evaluated to determine whether differences in LOS were influenced by demographic composition rather than race alone.

Length of stay (LOS) represents the number of inpatient

days associated with each hospitalization encounter. Two forms of LOS analysis were conducted in this study:

1. Total LOS: The cumulative number of hospitalization days for each racial group.
2. Average LOS: The mean hospitalization duration per patient within each racial group.

The distinction between total LOS and average LOS is important because total LOS may be strongly influenced by population size, whereas average LOS provides a standardized measure that better reflects individual patient experiences.

### 2.3. Data Quality and Missing Information

One notable feature of the dataset is the presence of missing race information represented by the “Null” category. This category reflects cases in which race was either unreported, unavailable, or improperly recorded during data collection. Missing demographic data present an important challenge in population health research because incomplete information can bias analyses and limit the accuracy of disparity assessments.

The “Null” category was not treated as a valid racial classification during comparative analysis because it may contain individuals from multiple unidentified populations. However, the category was retained within the dataset to maintain transparency regarding data completeness. The presence of missing race information highlights the importance of standardized demographic data collection practices in healthcare systems.

### 2.4. Population Characteristics

Analysis of the dataset revealed variation in demographic composition across racial groups. White patients represented the largest proportion of respiratory illness hospitalizations, particularly within older age categories. Black and Other racial groups demonstrated moderate representation across age groups, while Asian or Pacific Islander and American Indian populations represented smaller proportions of the hospitalized population.

The dataset also demonstrated a trend of increasing hospitalization frequency among older age groups across all racial categories. This pattern is consistent with public health evidence showing that respiratory illnesses disproportionately affect older adults due to increased comorbidity, weakened immune function, and greater disease severity.

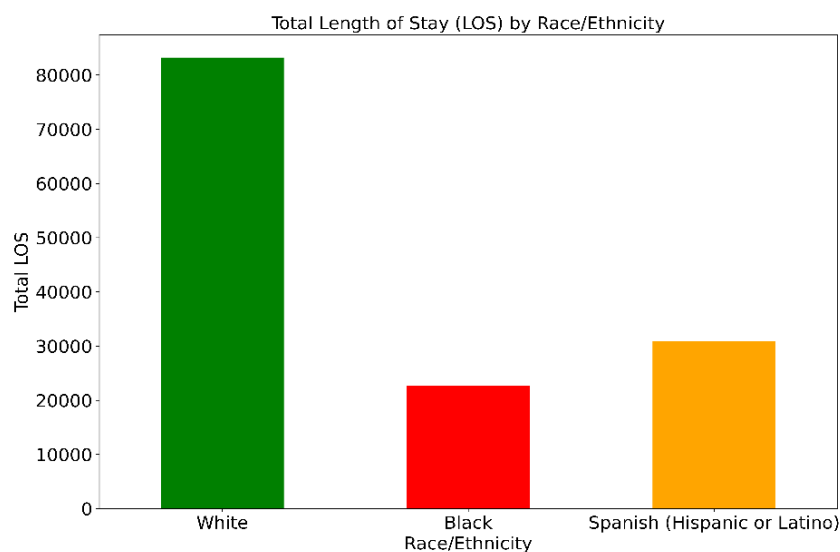
### 2.5. Relevance to Population Health Research

The Texas respiratory illness dataset provides valuable insight into healthcare utilization and demographic variation within hospitalized populations. By examining LOS across racial and age groups, the dataset supports investigation into potential healthcare disparities and factors influencing patient outcomes. The data also allow researchers to evaluate how demographic variables such as age and race interact to influence hospitalization patterns.

From a population health perspective, the dataset emphasizes the importance of adjusting for confounding variables and using standardized measures when interpreting disparities. Raw differences in hospitalization outcomes may not necessarily indicate inequity without considering population structure, social determinants of health, and healthcare access. Therefore, the dataset serves as a useful foundation for conducting evidence-based analyses aimed at improving healthcare equity and informing public health interventions within Texas healthcare systems.

### 3. Data Visualization

Figure 1 shows the total Length of Stay (LOS) by Race/Ethnicity among patients hospitalized for respiratory illness in Texas. The Figure illustrates the cumulative hospital length of stay across three demographic groups: White, Black, and Spanish (Hispanic or Latino) populations. White patients demonstrated the highest total LOS, exceeding 80,000 hospital days, while Spanish (Hispanic or Latino) patients accounted for approximately 30,000 hospital days and Black patients accounted for slightly more than 20,000 hospital days.



**Figure 1.** Total Length of Stay (LOS) by Race/Ethnicity among patients hospitalized for respiratory illness in Texas.

These differences in total LOS may reflect variations in population size, hospitalization frequency, age distribution, disease burden, healthcare utilization, and underlying demographic characteristics. Because total LOS represents aggregate hospitalization days rather than individual patient averages, the findings should not be interpreted as direct evidence of healthcare disparities without further adjustment for confounding variables such as age, comorbidities, socioeconomic status, and access to care. The visualization highlights the importance of considering both demographic composition and standardized outcome measures when evaluating population health patterns and respiratory illness burden across racial and ethnic groups in Texas.

Figure 2 shows an average length of stay (LOS) by Race/Ethnicity among patients hospitalized for respiratory illness in Texas. The Figure presents the mean hospital length of stay per patient across three demographic groups: White, Black, and Spanish (Hispanic or Latino) populations. Unlike the total LOS analysis, which reflects the cumulative burden of hospitalization across the entire population, average LOS standardizes the outcome at the individual patient level and therefore provides a more accurate comparison of hospitalization experience between demographic groups. The results

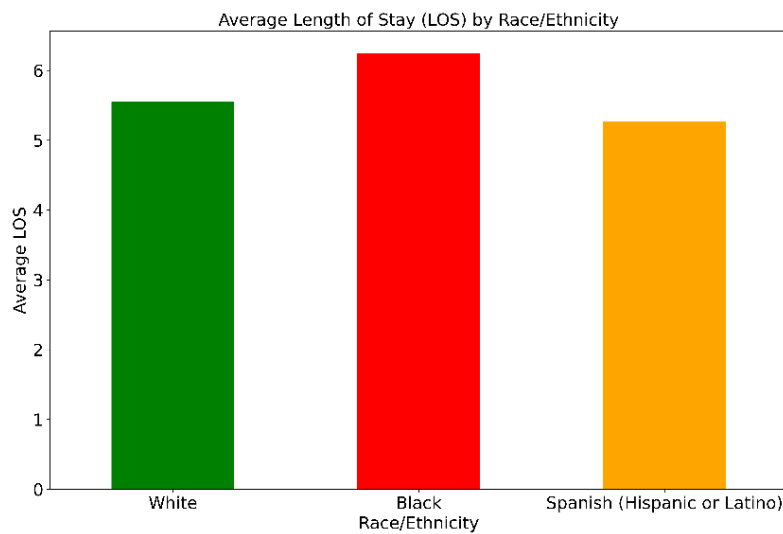
demonstrate that differences in average LOS across racial and ethnic groups are substantially smaller than differences observed in total LOS, suggesting that population size strongly influenced the cumulative hospitalization totals observed in the previous analysis.

The visualization indicates that White patients do not demonstrate elevated average LOS values relative to Black patients, while Spanish (Hispanic or Latino) patients exhibit intermediate hospitalization durations. However, the variation between groups is considerably narrower when measured using average LOS rather than total LOS. These findings suggest that raw cumulative hospitalization days may overstate apparent disparities when demographic population size is not considered. Average LOS provides a more standardized measure of healthcare utilization and may better reflect differences in patient-level disease severity, treatment duration, access to care, and clinical outcomes.

Interpretation of these findings should still be approached with caution because average LOS may be influenced by multiple confounding factors, including age distribution, comorbidities, socioeconomic status, insurance coverage, healthcare access, and severity of respiratory illness at admission. Older populations, for example, often require longer hospital stays

due to increased chronic disease burden and complications. Therefore, additional analyses using age-adjusted or multivariable statistical approaches would be necessary to determine

whether the observed differences represent true healthcare disparities or are primarily attributable to demographic and structural variation across populations.



**Figure 2.** The mean hospital length of stay per patient across three demographic groups: White, Black, and Spanish (Hispanic or Latino) populations.

### 4. Confounders for LOS

A mathematical formulation for confounders in your respiratory LOS analysis can be written using a regression model

$$LOS_i = \beta_0 + \beta_1 Race_i + \beta_2 Age_i + \beta_3 Sex_i + \beta_4 Insurance_i + \beta_5 Comorbidity_i + \beta_6 SocioeconomicStatus_i + \epsilon_i$$

Where:

- 1)  $LOS_i$  = Length of stay for patient i
- 2)  $\beta_0$  = Intercept
- 3)  $Race_i$  = Race/ethnicity category for patient i
- 4)  $Age_i$  = Age of patient iii
- 5)  $Sex_i$  = Biological sex/gender
- 6)  $Insurance_i$  = Insurance status
- 7)  $Comorbidity_i$  = Presence of chronic diseases or severity indicators
- 8)  $SocioeconomicStatus_i$  = Income, education, or neighborhood indicators
- 9)  $\epsilon_i$  = Random error term

The coefficients:  $\beta_i$  measure the independent contribution of each variable to LOS after adjusting for the others.

Let us consider the following confounders for LOS:

where Length of Stay (LOS) is the outcome variable and race/ethnicity is evaluated while controlling for other influencing variables (confounders).

In our study, a general multiple linear regression model would look like this de Melo [21]:

Confounder	Description
Sex	Respiratory outcomes may differ biologically
Insurance status	Influences healthcare access
Comorbidities	Diabetes, COPD, asthma increase LOS
Severity of illness	Sicker patients require longer care
Socioeconomic status	Affects healthcare utilization
Smoking history	Strong respiratory risk factor

**Table 1.** LOS dependent confounders.

Confounder	Description
Age	Older patients generally stay longer

Table 1 shows the relationship between race and LOS changes after adjustment for confounding variables. To account for potential confounding effects, a multivariable regression framework may be applied in which length of stay (LOS) serves as the dependent variable and race/ethnicity represents the primary independent variable. Additional covariates including age, sex, comorbidities, insurance status, and socioeconomic factors may be incorporated into the model to control for demographic and clinical variation influencing

hospitalization outcomes. This approach allows assessment of whether observed differences in LOS remain significant after adjustment for relevant confounding variables.

### 4.1. Age as a Confounder

Figure 3 shows an average hospital length of stay (LOS) across major age groups stratified by race and ethnicity. The grouped bar chart displays the mean LOS in days for White patients (green), Black patients (red), and Hispanic/Latino patients (orange) within each age category: 0 to 17 years, 18 to 44 years, 45 to 64 years, 65 to 74 years, and 75 years and older. Numerical labels above each bar indicate the average LOS for the corresponding subgroup.

The Figure demonstrates substantial variation in LOS across both age and racial/ethnic categories, illustrating the potential confounding effect of age in analyses of healthcare disparities. LOS generally increases with advancing age, consistent with the greater clinical complexity, comorbidity burden, and healthcare utilization typically observed among older hospitalized patients. Because age is strongly associated with the outcome (LOS), differences in age composition across racial or ethnic groups can distort crude comparisons of hospitalization duration.

The markedly elevated LOS observed among Black patients aged 0 to 17 years suggests the presence of subgroup-

specific clinical or utilization patterns that substantially influence overall disparity estimates. In contrast, LOS differences among adult age groups are smaller and demonstrate different directional patterns across racial and ethnic categories. White patients exhibit progressively increasing LOS across older age strata, while Hispanic/Latino patients demonstrate intermediate LOS values across most age groups. These heterogeneous patterns indicate that the association between race/ethnicity and LOS is not uniform across age strata.

From an epidemiologic perspective, the Figure provides visual evidence that age may function as a confounder because it is associated with both the exposure (race/ethnicity) and the outcome (LOS). If one racial or ethnic group contains a larger proportion of older individuals, crude estimates of racial disparity may partially reflect underlying demographic age structure rather than an independent effect of race or ethnicity on hospitalization outcomes. Therefore, statistical adjustment for age is necessary to obtain unbiased estimates of racial and ethnic disparities in LOS.

The stratified age-group presentation shown in this Figure helps distinguish whether observed disparities are attributable primarily to demographic differences in population structure or whether disparities persist within comparable age strata after adjustment. Persistent differences across age groups after adjustment would support the presence of independent healthcare disparities beyond age-related effects alone.

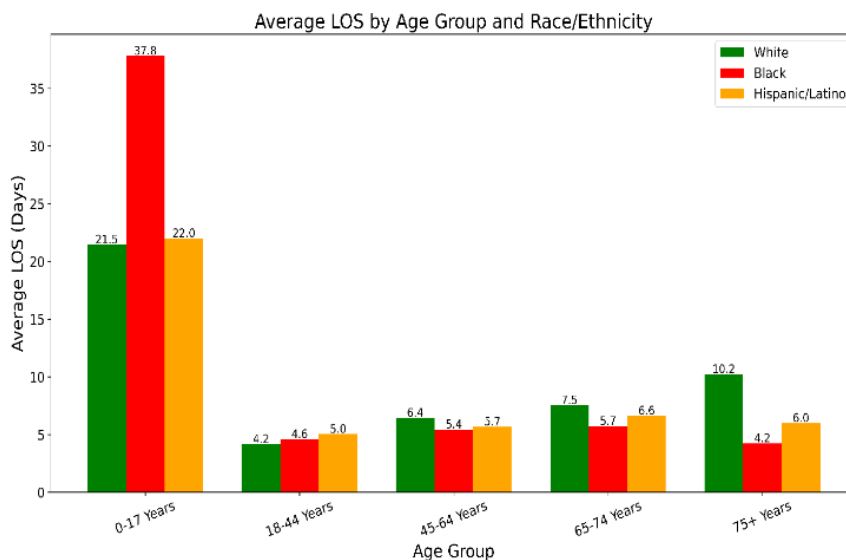


Figure 3. Average hospital length of stay (LOS) across major age groups stratified by race and ethnicity. Among white patients, average LOS increases steadily with age. Among Black and Hispanic/Latino patients, LOS appears flatter or even lower in older age groups. Black children (0 to 17 years) have an unusually high average LOS compared with all other groups.

Table 2. LOS summary.

Group	Number	Female%	Mean LOS	Median LOS	SD LOS
Black	3278	55.3	6.32	4.0	9.53

Group	Number	Female%	Mean LOS	Median LOS	SD LOS
Hispanic/Latino	5531	51.7	5.20	3.0	6.63
White	14043	55.1	5.48	4.0	6.75

Table 2 presents descriptive statistics of hospital length of stay (LOS) across major racial and ethnic groups in the respiratory cohort. White patients constituted the largest cohort (14,043 patients), followed by Hispanic/Latino (5,531) and Black patients (3,278). Female representation was relatively similar across groups, ranging from 51.7% to 55.3%. Black patients demonstrated the highest mean LOS (6.32 days) and the largest variability in hospitalization duration (SD = 9.53),

suggesting a broader distribution of disease severity, comorbidities, or healthcare utilization patterns within this population. Hispanic/Latino patients showed the lowest mean and median LOS values, while White patients exhibited intermediate LOS measures. The consistently lower median LOS compared with mean LOS in all groups indicates a right-skewed distribution, where a smaller subset of patients experienced substantially prolonged hospitalizations, creating a “long-tail” effect often observed in respiratory care datasets.

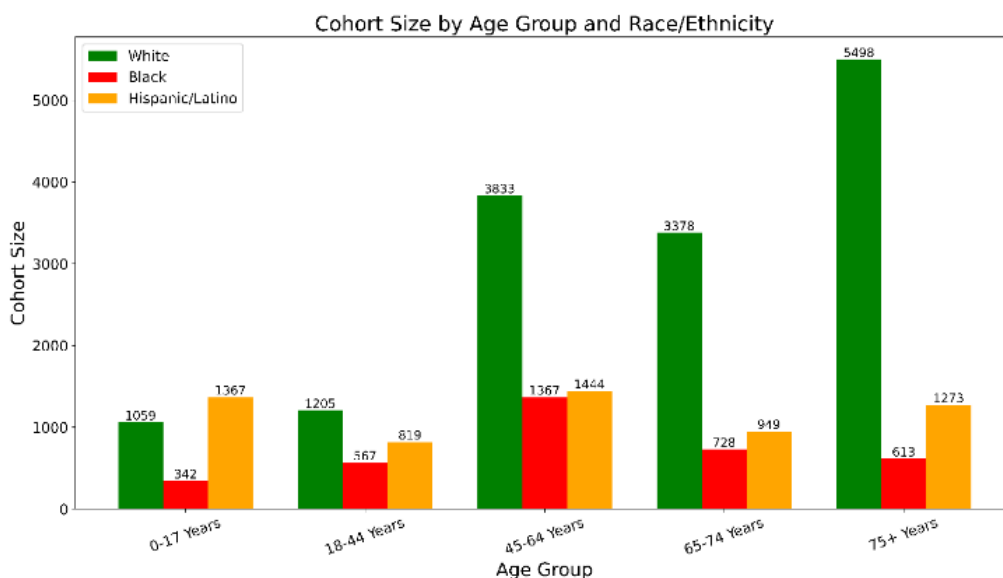


Figure 4. The cohort size distribution across major age groups stratified by race and ethnicity among patients hospitalized with respiratory conditions.

Figure 4 shows the cohort size distribution across major age groups stratified by race and ethnicity among patients hospitalized with respiratory conditions. The grouped bar chart displays the number of patients within each age category for White patients (green), Black patients (red), and Hispanic/Latino patients (orange). Numerical labels above each bar represent the total cohort size for the corresponding subgroup. Age categories were consolidated into five clinically meaningful strata: 0 to 17 years, 18 to 44 years, 45 to 64 years, 65 to 74 years, and 75 years and older.

The Figure demonstrates substantial differences in demographic composition across racial and ethnic groups, with White patients comprising the largest cohort across all adult age categories, particularly among older adults aged 75 years and above. In contrast, Black and Hispanic/Latino cohorts are

smaller overall and exhibit different age distributions. Hispanic/Latino patients show relatively greater representation within younger age groups, whereas White patients demonstrate marked predominance in older age strata. Black patient cohort sizes are intermediate but decline substantially in the oldest age category.

These differences in age composition are epidemiologically important because age is strongly associated with hospital length of stay (LOS), disease severity, comorbidity burden, and healthcare utilization. Consequently, crude comparisons of LOS across racial and ethnic groups may be substantially influenced by underlying demographic structure rather than reflecting independent racial or ethnic disparities alone. For example, the disproportionately large number of older White patients may contribute to higher overall LOS estimates in

White cohorts because advanced age is typically associated with prolonged hospitalization.

The Figure therefore provides important contextual evidence supporting age as a potential confounder in analyses of racial and ethnic disparities in LOS. A confounder is a variable associated with both the exposure (race/ethnicity) and the outcome (LOS) that can distort observed associations if not properly controlled. The markedly uneven cohort distributions observed across age strata indicate that statistical adjustment for age is necessary to obtain unbiased estimates of racial and ethnic differences in hospitalization outcomes.

In addition, the relatively small cohort sizes observed in certain subgroups, particularly among Black patients aged 75 years and older, suggest that subgroup-specific LOS estimates may be sensitive to statistical instability and outlier effects. Small cohort sizes can produce exaggerated mean LOS values when a limited number of patients experience unusually prolonged hospitalizations. Therefore, interpretation of age-stratified LOS differences should incorporate consideration of cohort size, variability, and distributional characteristics in addition to crude mean estimates alone.

## 4.2. Disease as a Confounder

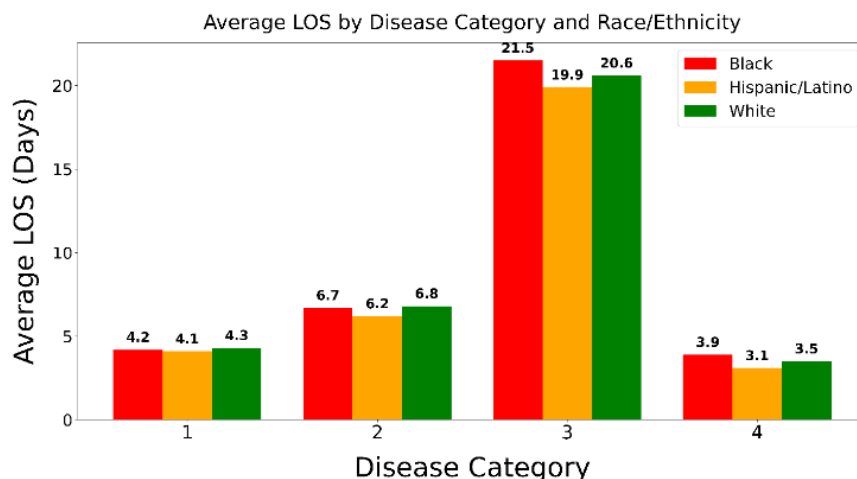
Substantial variation in LOS was observed across respiratory disease categories. Patients requiring prolonged ventilator support demonstrated the highest LOS across all racial and

ethnic groups, with average hospitalization durations exceeding 19 days. In contrast, viral illness and uncomplicated pneumonia were associated with substantially shorter hospital stays.

Figure 5 shows that across disease categories, Black patients demonstrate the longest LOS, followed by White and Hispanic/Latino patients. For example, within the ventilator-support category, mean LOS was 21.5 days among Black patients, 20.6 days among White patients, and 19.9 days among Hispanic/Latino patients. Similar patterns were observed in pulmonary edema and respiratory failure, where Black patients exhibited higher average LOS compared with other groups.

Visualization of LOS across age strata demonstrated a strong age-dependent increase in hospitalization duration. Older patients consistently exhibited longer and more variable LOS distributions, particularly among individuals aged 65 years and older. Adjustment for age attenuated several observed racial and ethnic differences in LOS, indicating that demographic age structure contributed substantially to crude disparity estimates.

Age fulfilled the epidemiologic criteria for a confounder because it was independently associated with both race/ethnicity and LOS while not representing an intermediate causal variable. Interaction analyses further suggested that the magnitude of racial disparities may differ across age groups, indicating possible effect modification in older populations.



**Figure 5.** The cohort size distribution across major groups stratified by race and disease category hospitalized with respiratory conditions. “1” refers to Pneumonia, “2” refers to Pulmonary Edema and Respiratory Failure, “3” refers to Respiratory Disease with 96+ hours of ventilator support, and finally “4” corresponds to Viral illness.

## 5. Discussion

This study evaluated racial and ethnic differences in hospital length of stay (LOS) among patients hospitalized with respiratory illnesses and examined the extent to which age acted

as a confounding variable in disparity estimation. Several important findings emerged from the analysis. First, LOS varied substantially across respiratory disease categories, with the longest hospitalization durations observed among patients requiring prolonged ventilatory support. Second, crude racial and ethnic differences in LOS were identified across multiple disease categories, with Black patients generally demonstrating

the highest average LOS. Third, and most importantly, adjustment for age attenuated several observed disparities, indicating that age distribution substantially influenced crude estimates of racial and ethnic differences in hospitalization outcomes.

The observed increase in LOS among older patients was clinically expected and epidemiologically significant. Advancing age is associated with greater physiologic vulnerability, multimorbidity, reduced pulmonary reserve, frailty, and increased risk of complications during hospitalization. Older adults hospitalized with respiratory illnesses frequently require more intensive monitoring, prolonged oxygen therapy, rehabilitation services, and extended discharge planning, all of which contribute to longer hospitalization durations. The visualization analyses demonstrated a progressive upward shift in LOS distributions across age categories, with widening interquartile ranges and greater variability among elderly patients. These findings reinforce the strong relationship between age and inpatient healthcare utilization.

The present findings demonstrate why age represents a critical confounding variable when evaluating racial and ethnic disparities in LOS. Confounding occurs when a third variable is independently associated with both the exposure and the outcome, thereby distorting the apparent association between them. In this analysis, age satisfied the major epidemiologic criteria for confounding because it was strongly associated with LOS and differed across racial and ethnic populations. As a result, crude LOS comparisons may partially reflect demographic age structure rather than independent effects attributable to race or ethnicity alone.

After adjustment for age, several racial and ethnic differences in LOS decreased substantially, suggesting that a meaningful proportion of the crude disparities was attributable to differences in underlying population age distributions. This finding has important methodological implications. Studies that fail to account for demographic confounding may overestimate healthcare disparities and potentially generate misleading interpretations regarding inequities in clinical outcomes. The results therefore emphasize the importance of multivariable adjustment and stratified analysis when examining healthcare utilization measures such as LOS.

Despite attenuation after adjustment, some racial differences persisted, particularly among Black patients in severe respiratory disease categories. Several explanations may contribute to these residual disparities. Differences in disease severity at presentation, access to outpatient preventive care, socioeconomic conditions, chronic comorbidity burden, healthcare access, delayed presentation, and structural inequities may all influence hospitalization duration. In addition, social determinants of health such as housing instability, transportation limitations, caregiver availability, and barriers to post-discharge follow-up may prolong discharge readiness and contribute to extended LOS in vulnerable populations.

The highest LOS values were observed among patients with respiratory system diagnoses requiring ventilator support for  $\geq 96$  hours. This finding likely reflects the substantial clinical

complexity associated with prolonged mechanical ventilation, including critical illness, respiratory failure, intensive care utilization, secondary infections, prolonged rehabilitation, and delayed recovery trajectories. Importantly, although all racial and ethnic groups demonstrated markedly elevated LOS within this category, Black patients exhibited the highest mean hospitalization duration. This observation suggests that disparities may become more pronounced among patients with severe disease burden and intensive resource utilization.

The study also demonstrated the value of visualization techniques in epidemiologic disparity research. Stratified boxplots and grouped bar plots provided intuitive visual evidence of age-related LOS gradients and highlighted the impact of demographic composition on crude disparity estimates. Visualization of LOS distributions across age strata allowed clearer identification of skewness, variability, and extreme outlier behavior that may not be fully captured by summary statistics alone. Such graphical approaches improve interpretability and enhance communication of complex confounding relationships in healthcare disparities research.

Several limitations should be acknowledged. First, the analysis was retrospective and observational, limiting causal inference. Second, LOS may be influenced by numerous unmeasured confounders not included in the present analysis, including comorbidity burden, insurance status, socioeconomic status, hospital characteristics, intensive care utilization, severity indices, and discharge disposition. Third, categorization of race and ethnicity may not fully capture social and cultural heterogeneity within populations. Fourth, LOS distributions demonstrated substantial skewness and outlier behavior, which may affect regression assumptions despite adjustment efforts.

Nevertheless, the study has several strengths. The analysis incorporated multiple respiratory disease categories, detailed age stratification, visualization-based epidemiologic assessment, and multivariable modeling approaches. The use of publication-quality graphical methods improved interpretability of confounding effects and facilitated clearer understanding of demographic influences on LOS disparities.

In conclusion, age is an important confounding factor in the estimation of racial and ethnic disparities in hospital LOS among patients with respiratory illnesses. Much of the observed crude variation in LOS across racial and ethnic groups reflects differences in underlying age distributions rather than entirely independent disparity effects. Appropriate adjustment for demographic variables is therefore essential for accurate interpretation of healthcare disparities. Future studies should incorporate broader social determinants of health, disease severity metrics, and longitudinal outcomes to further clarify the complex mechanisms underlying racial and ethnic variation in hospitalization outcomes.

## 6. Conclusion

This study demonstrated significant racial and ethnic differences in hospital length of stay among patients hospitalized

with respiratory illnesses, particularly in severe disease categories requiring prolonged ventilatory support. However, the analysis also showed that age substantially influenced these observed disparities and functioned as an important confounding variable in LOS estimation. Adjustment for age attenuated several crude racial and ethnic differences, indicating that underlying demographic age structure contributed meaningfully to variation in hospitalization duration. These findings emphasize the necessity of incorporating demographic adjustment and stratified epidemiologic analysis when evaluating healthcare disparities. Accurate interpretation of LOS outcomes requires consideration of both clinical severity and population characteristics to avoid overestimation of race-associated effects. Future investigations should incorporate broader social determinants of health, comorbidity burden, and longitudinal outcome measures to further clarify the mechanisms contributing to disparities in respiratory hospitalization outcomes.

## Abbreviations

CDC	Centers for Disease Control and Prevention
COPD	Chronic Obstructive Pulmonary Disease
LOS	Length of Stay
SD	Standard Deviation

## Author Contributions

**Philip de Melo:** Conceptualization, Software  
**Mildred Jackson:** Methodology  
**Aleysia James:** Investigation  
**Keanna Lawrence:** Validation  
**Shakira McElveen:** Visualization  
**Heaven Mitchell:** Writing – review & editing  
**Angel Tolbert-Abdullah:** Visualization  
**Ashley Wallace:** Resources  
**Jasmine Waller:** Writing – review & editing

## Conflicts of Interest

The authors declare no conflicts of interest.

## References

- [1] Nash, D. B., Skoufalos, A., Fabius, R. J., & Oglesby, W. H. (2019). *Population health: Creating a culture of wellness* (3rd ed.). Jones & Bartlett Learning.
- [2] Kliethermes, S. A., Asif, I. M., Blauwet, C., Christensen, L., Coleman, N., Lavalley, M. E., Moeller, J. L., Phillips, S. F., Rao, A., Rizzone, K. H., Sund, S., Tanji, J. L., Tuakli-Wosornu, Y. A., & Stafford, C. D., II. (2024). Focus areas and methodological characteristics of North American-based health disparity research in sports medicine: A scoping review. *British Journal of Sports Medicine*, 58(3), 164–171. <https://doi.org/10.1136/bjsports-2023-107607>
- [3] Baker, C., Chinaka, O., & Stewart, E. C. (2021). The time is now: Why we must identify and address health disparities in sport and recreation injury. *Injury Epidemiology*, 8(1), 25. <https://doi.org/10.1186/s40621-021-00318-4>
- [4] Tucker, C., Antoniak, K., & Edison, B. (2021). Athlete medicine applied in a pandemic: Disparities among athletes and performers reveal the need for a true “revolution” in health care. *BMJ Open Sport & Exercise Medicine*, 7(1), e000977. <https://doi.org/10.1136/bmjsem-2020-000977>
- [5] Tuakli-Wosornu, Y. A., Zondi, P. C., Knudson, G., et al. (2021). Sports equity: A new BJSM e-Edition brings the fundamentals back into focus. *British Journal of Sports Medicine*, 55(24), 1385–1386. <https://doi.org/10.1136/bjsports-2021-104878>
- [6] Puka, K., Kilian, C., Zhu, Y., et al. (2023). Can lifestyle factors explain racial and ethnic inequalities in all-cause mortality among US adults? *BMC Public Health*, 23(1), 1591. <https://doi.org/10.1186/s12889-023-16444-9>
- [7] Brown, C. S., Masters, K. S., & Huebschmann, A. G. (2018). Identifying motives of midlife Black triathlete women using survey transformation to guide qualitative inquiry. *Journal of Cross-Cultural Gerontology*, 33(1), 1–20. <https://doi.org/10.1007/s10823-017-9341-3>
- [8] Calzo, J. P., Roberts, A. L., Corliss, H. L., Blood, E. A., Kroshus, E., & Austin, S. B. (2014). Physical activity disparities in heterosexual and sexual minority youth ages 12–22 years old: Roles of childhood gender nonconformity and athletic self-esteem. *Annals of Behavioral Medicine*, 47(1), 17–27. <https://doi.org/10.1007/s12160-013-9570-y>
- [9] Fagher, K., DeLuca, S., Derman, W., & Blauwet, C. (2023). Optimising health equity through para sport. *British Journal of Sports Medicine*, 57(3), 131–132. <https://doi.org/10.1136/bjsports-2022-106653>
- [10] Baria, M. R., Terry, M. J., Driscoll, S. W., Andrews, K. L., Soma, D. B., & Prideaux, C. C. (2015). Wrestlers with limb deficiencies: A descriptive study. *American Journal of Physical Medicine & Rehabilitation*, 94(12), 1052–1057. <https://doi.org/10.1097/PHM.0000000000000299>
- [11] Barfield, J. P., & Malone, L. A. (2013). Perceived exercise benefits and barriers among power wheelchair soccer players. *Journal of Rehabilitation Research & Development*, 50(2), 231–238. <https://doi.org/10.1682/JRRD.2011.09.0170>
- [12] Braza, D. W., Iverson, M., Lee, K., Hennessy, C., & Nelson, D. (2018). Promoting physical activity by creating awareness of adaptive sports and recreation opportunities: An academic-community partnership perspective. *Progress in Community Health Partnerships*, 12(2), 165–172. <https://doi.org/10.1353/cpr.2018.0032>
- [13] Budavari, A. C., Pas, E. T., Azad, G. F., & Volk, H. E. (2022). Sitting on the sidelines: Disparities in social, recreational, and community participation among adolescents with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 52(8), 3399–3412. <https://doi.org/10.1007/s10803-021-05206-6>

- [14] Hartill, M., & Rulofs, B. (2021). Child abuse in sport: European statistics (CASES). The prevalence and characteristics of interpersonal violence against children within and outside sport in six European countries. Presented at the Transfer-Symposium, World Athletics, Monaco.
- [15] Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, 8(1), 19–32.  
<https://doi.org/10.1080/1364557032000119616>
- [16] Peters, M. D. J., Marnie, C., Tricco, A. C., et al. (2021). Updated methodological guidance for the conduct of scoping reviews. *JBIEvidence Implementation*, 19(1), 3–10.  
<https://doi.org/10.1097/XEB.0000000000000277>
- [17] Tricco, A. C., Lillie, E., Zarin, W., et al. (2018). PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, 169(7), 467–473.  
<https://doi.org/10.7326/M18-0850>
- [18] Bramer, W. M., Giustini, D., de Jonge, G. B., Holland, L., & Bekhuis, T. (2016). De-duplication of database search results for systematic reviews in EndNote. *Journal of the Medical Library Association*, 104(3), 240–243.  
<https://doi.org/10.3163/1536-5050.104.3.014>
- [19] De Melo, P., DiLella, M., Holman, T., & McElveen, S. (2025). Accurate prediction of survival based on Kaplan-Meier analytics. *Cancer Research Journal*, 13(4), 173–185.  
<https://doi.org/10.11648/j.crj.20251304.14>
- [20] De Melo, P., & St. Rose, M. (2025). Accurate classification of diabetes via PM Generative AI. *Advances in Bioscience and Biotechnology*, 16, 379–409.  
<https://doi.org/10.4236/abb.2025.169025>
- [21] De Melo, P. (2024). *Public health informatics and technology*. Library of Congress. ISBN 979-8894090962.