Forecasting Cataract Prevalence in the United States Based on Race: A Data-driven Analysis for 2030, 2040, and 2050

Azra Emekci[1], Shreya Kochar, Aneesha Sreerama, Andrew Wang

Abstract

The number of people diagnosed with cataracts in the United States is increasing each year. Utilizing data-driven analysis, the study will examine existing cataract statistics and demographic data from previous years to develop reliable projections. The research will employ advanced statistical modeling and forecasting techniques to provide insights into the potential burden of cataracts on specific races in the coming decades. This research seeks to provide valuable insights into the patterns of cataract prevalence in multiple races and identify potential risk factors.

1. Introduction

Cataracts, a common age-related ocular condition, have been on the rise in the United States, with the national prevalence rate increasing by over 2% from 2014 to 2019. In those 6 years, cataract prevalence in Asians has increased from 30.40% to 32.93% and from 23.50% to 25.91% in the non- Hispanic Black population. Similarly, the cataract prevalence increased by 1.49% in the Hispanic population in America. Following an identical trend, there was an increase from 33.70% cataract prevalence to 35.87% in the non-Hispanic White population.[1]

Cataracts are characterized by the clouding of the eye's lens which causes visual impairment and often leads to blindness if left untreated. Cataracts can significantly impact daily activities such as reading, driving, and recognizing faces. By removing cataracts, individuals can regain their ability to perform these activities and enhance their overall quality of life. One of the pioneering studies in this field was conducted by the Eye Diseases Prevalence Research Group in 2004. This study aimed to determine the prevalence of cataract in the United States and projected changes in prevalence figures up to the year 2020. By analyzing data from major population-based studies conducted in the United States, Australia, Barbados, and Western Europe, the researchers provided summary prevalence estimates for cataract, stratified by race/ethnicity, age, and gender.

The findings revealed an anticipated increase in cataract cases over the 20- year period. Cataracts were identified as the leading cause of vision loss in the United States, emphasizing the need for a better understanding of the prevalence rates among different racial groups. Notably, significant disparities in cataract prevalence were observed between various racial groups, with women exhibiting higher age-adjusted prevalence than men in the United States.

Looking ahead, the study projected a substantial rise in the total number of Americans affected by cataracts by the year 2020, signifying a 50% increase over two decades. This projection highlights the importance of continued research and interventions to address the growing burden of cataracts on public health.[2]

To further explore the prevalence of cataracts within specific racial and ethnic communities, subsequent studies have investigated their prevalence in targeted populations. For instance, the Los Angeles Latino Eye Study (LELAS) conducted in 2009 focused on the prevalence of cataracts within the Latino population residing in Los Angeles. The study revealed a cataract prevalence rate of 5.6% among participants of Hispanic descent aged 40 years and older. While this percentage may appear modest, given the estimated population size of elderly Latinos in the United States, the absolute number of individuals affected by cataracts remains substantial.

Additionally, research has been conducted to examine racial disparities in the prevalence of cataracts and access to necessary surgical interventions. A study conducted in Florida using data from the National Institute of Eye Health and the Healthcare Cost and Utilization Project revealed notable differences in cataract prevalence between white and African-American populations. White individuals exhibited a higher prevalence of cataracts, indicating potential disparities in the occurrence of this ocular condition among different racial groups.[3]

While previous studies have examined cataract prevalence and its associated risk factors, few have employed data science techniques to predict the cataract prevalence. Forecast predictions are essential because they provide estimates of future values based on historical data and patterns. From the data provided by Vision and Eye Health Surveillance System (VEHSS) from Centers for Disease Control and Prevention (CDC) which "represent[s] the primary surveillance measures of cumulative

[1] CDC. "Vision and Eye Health Surveillance System (VEHSS)." Centers for Disease Control and Prevention, Centers for Disease Control and Prevention

[2] The Eye Diseases Prevalence Research Group*. "Prevalence of Cataract and Pseudophakia/Aphakia among Adults in the United States." Archives of Ophthalmology, JAMA Network

[3] Richter, Grace M, et al. "Prevalence of Visually Significant Cataract and Factors Associated with Unmet Need for Cataract Surgery: Los Angeles Latino Eye Study." Ophthalmology, U.S. National Library of Medicine

diagnosed and undiagnosed disorders," the central problem of future cataract trajectory was addressed in this interdisciplinary research project. Their data set "Annual prevalence of diagnosed cataracts" was employed to understand how the prevalence of diagnosed cataracts might evolve in the coming years using data science and programming.[4]

2. Background

Despite its impact, comprehensive and nationally representative estimates of cataract prevalence among various racial groups have been limited. In order to address this knowledge gap, a seminal study was conducted by the Eye Diseases Prevalence Research Group in 2004. This research aimed to determine the prevalence of cataract and pseudophakia/aphakia in the United States and project the expected changes in these prevalence figures by the year 2020. By analyzing data from major population-based studies conducted in the United States, Australia, Barbados, and Western Europe, the researchers prepared summary prevalence estimates for cataract and pseudophakia/aphakia separately for black, white, and Hispanic persons, taking age, gender, and race/ethnicity into account. The study's findings highlighted the anticipated increase in cataract cases over 20 years. As one of the first comprehensive estimates of cataract prevalence in the United States, this study serves as a crucial foundation for further research. The study revealed that an estimated 20.5 million Americans older than 40 years (17.2% prevalence) had cataracts in either eye at the time of the study. Cataract was identified as the leading cause of vision loss in the United States. Significant differences in cataract prevalence were observed among different racial groups, with women having a significantly higher age-adjusted prevalence of cataracts than men in the United States. The prevalence of cataracts increased with age for both black and white persons. Looking ahead, the study projected that the total number of Americans with cataracts would rise to approximately 30.1 million by the year 2020, signifying an expected increase of about 50% in the next two decades as the U.S. population continues to age. The study's comparison of cataract prevalence data from the United States, Australia, and Western Europe found generally similar rates across diverse studies of white individuals.[5]

A study conducted by the Los Angeles Latino Eye Study in 2009 focused specifically on investigating the prevalence of cataracts within the Latino population residing in Los Angeles. To accomplish this, the researchers extended invitations to 6142 participants of Hispanic descent, all aged 40 years or older, to undergo a comprehensive eye examination. Through meticulous examination and data analysis, the study revealed that 344 out of the 6142 participants were afflicted with a significant eye cataract in one or both eyes or had previously undergone cataract surgery, resulting in a prevalence rate of 5.6%. While this percentage may seem relatively small at first glance, the implications of cataract prevalence among the Latino community in the US are profound. Considering the estimated population size of the elderly Latino community in the country, even this seemingly modest prevalence rate translates to a considerable number of individuals affected by cataracts.[6]

In 2010, a study was conducted in Florida using data from the National Institute of Eye Health and the Healthcare Cost and Utilization Project. The study focused on cataract cases and procedures in four race/gender groups (African-American male and female, white male and female) aged 65 and over. The study aimed to examine the association between racial composition and the disparities in receiving necessary cataract surgery. According to their data, 1,116,081 of Florida's 2,678,014 elderly white population has cataracts, resulting in about 42% prevalence. Similarly, of the 275,968 elderly African-Americans in Florida, 91,290 had cataracts, a 33% prevalence. Their findings show that White individuals exhibited a higher prevalence of cataracts by approximately 10%.[7]

A similar pattern was observed when analyzing the National records from CDC. There seems to be a 10% difference between cataracts prevalence in whites and African-Americans in every year of the data, from 2014 to 2019. For example, the cataract prevalence nationally of African-Americans in 2014 was 23.5% and whites was 33.7%.

3. Methodology

The GitHub link below includes the code and the dataset used in this research. The code is located in Time_Series_Forecast_of_Cataracts_by_Race.ipynb and the dataset is in Dataset.csv.

`https://github.com/AzraEm/forecastingCataractPrevalence`

As the world's population continues to age, the prevalence of cataracts is expected to rise in the coming decades. To gain insights into the future prevalence of cataracts, a custom forecast model was developed using a linear regression algorithm in Python. This model utilized the nationwide cataract prevalence data from the years 2014 to 2019 collected by Vision & Eye Health Surveillance System (VEHSS) from the Centers for Disease Control and Prevention (CDC) by Medicare. It was filtered from a larger dataset containing the national cataract prevalence comparing all years, from 2014 to 2019.

The Pandas library in Python was employed to read, write, and process the dataset stored in the CSV file. Pandas allows for efficient data manipulation and preparation, ensuring the dataset is suitable for analysis. The linear regression algorithm from Scikit-learn was chosen to develop the custom forecast model. Linear regression is a powerful technique for predicting numerical values based on historical data trends. By using the cataract

[4]"Vision and Eye Health Surveillance System (VEHSS)." Centers for Disease Control and Prevention, Centers for Disease Control and Prevention

[5]The Eye Diseases Prevalence Research Group*. "Prevalence of Cataract and Pseudophakia/Aphakia among Adults in the United States." Archives of Ophthalmology, JAMA Network

[6]Richter, Grace M, et al. "Prevalence of Visually Significant Cataract and Factors Associated with Unmet Need for Cataract Surgery: Los Angeles Latino Eye Study." Ophthalmology, U.S. National Library of Medicine

[7]https://journals.plos.org/plosone/article? id=10.1371/journal.pone.0142459#abstract0

prevalence figures from 2014 to 2019, the model was trained to identify patterns and relationships between the years and the corresponding prevalence rates.

Using the Scikit-learn Linear Regression tools, the model analyzed the historical data and projected the future prevalence of cataracts for the years 2030, 2040, and 2050. The model assumes that the historical trends observed over the given years will continue in a linear fashion. However, it's important to note that the accuracy of the forecast relies heavily on the reliability and representativeness of the data.

To provide a clear and visual representation of the forecast results, the Matplotlib library (imported as plt) was utilized. Matplotlib.pyplot allows for the creation of various graphs and visualizations, enabling easy interpretation of the predicted trends. Visualizations such as line plots, bar graphs, and area charts can help stakeholders understand the expected rise in cataract prevalence over the years.

4. Results

Year	Asian	non-Hispanic Black	Hispanic	non-Hispanic White
2030	38.94	31.06	28.2	40.62
2040	44.24	35.66	31.2	44.92
2050	49.54	40.26	34.2	49.22

Figure 1: Predicted Cataract Prevalence of Each Race.

Based on the projected future cataract prevalence data, it is evident that the prevalence of cataracts is expected to increase across all racial and ethnic groups in the coming years, as shown in Figure 1. The data indicates that by the year 2030, the prevalence rates are estimated to be approximately 38.94% for Asians, 31.06% for non-Hispanic Black individuals, 28.2% for Hispanics, and 40.62% for non-Hispanic White individuals. These values are projected to further rise by 2040 and 2050. By 2050, the prevalence rates are projected to reach around 49.54% for Asians, 40.26% for non-Hispanic Black individuals, 34.2% for Hispanics, and 49.22% for non-Hispanic White individuals. The forecasted trends in America suggest a general upward trajectory of cataract prevalence across all racial and ethnic groups consistent with the aging population and aligns with the upward trend observed in the historical data.

Figure 2 illustrates the rate of increase for each race in percent (%) from 2019 to the years the cataract prevalence was predicted, 2030, 2040, and 2050.

Interval	Asian	non-Hispanic Black	Hispanic	non-Hispanic White
2019-2030	18.25 %	19.88 %	13.76 %	13.24 %
2019-2040	34.34 %	37.63 %	25.86 %	25.23 %
2019-2050	50.44 %	55.38 %	37.96 %	37.22 %

Figure 2: Rate of Increase for Each Race in the Given Intervals.

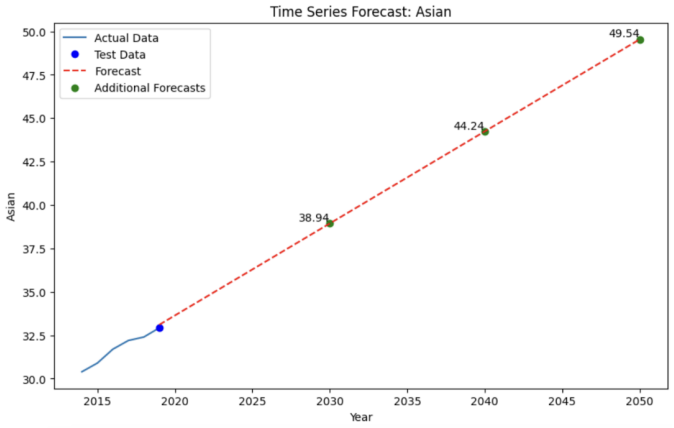

Figure 3: Cataract Prevalence for Asians.

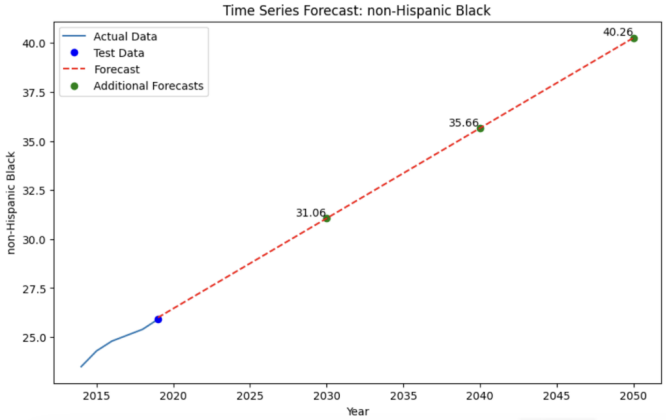

Figure 4: Cataract Prevalence for non-Hispanic Blacks.

Above, it is seen that non-Hispanic Blacks have the highest rate of increase every interval despite the fact that they have the second lowest cataract prevalence in the nation. In contrast, the rate of increase in the Asian population is the second highest when it was predicted to have the second highest cataract prevalence in 2030 and 2040, and in 2050, it would have the highest cataract prevalence. The rate of increase is very similar in Hispanics and non-Hispanic Whites throughout the given intervals, with about 0.63% difference, even though their predicted prevalence rate for each year has a difference of over 10%. It's observed that cataracts seem to affect every race differently.

The following graphs generated by the model plots the cataract prevalence in years 2014, 2015, 2016, 2017, 2018, and 2019 in blue and the predicted prevalence of years 2030, 2040, and 2050 in green with the red trend line. The data was graphed by race. Figure 3 is for Asian, Figure 4 is for non-Hispanic Black, Figure 5 is for Hispanic, and Figure 6 is for non-Hispanic Whites.

5. Summary and conclusions

The research paper addresses the rising prevalence of cataracts in the United States and its potential impact on differ-

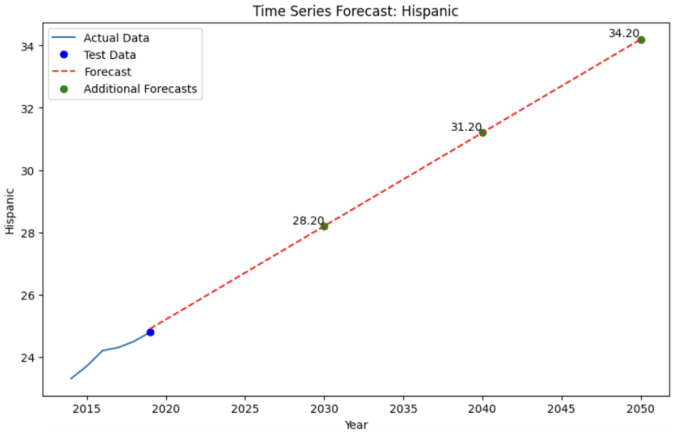

Figure 5: Cataract Prevalence for Hispanics.

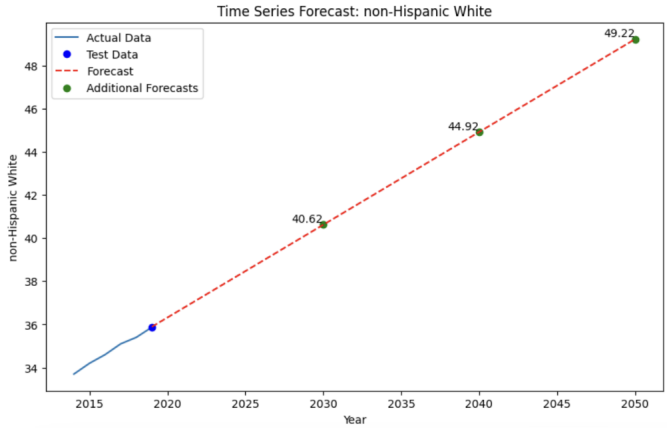

Figure 6: Cataract Prevalence for non-Hispanic Whites.

ent racial groups. The study employs advanced statistical modeling and forecasting techniques to project cataract prevalence trends up to 2050, providing insights into potential disparities and future challenges.

The introduction highlights the increasing prevalence of cataracts across racial groups in recent years, particularly among Asians, non-Hispanic Blacks, Hispanics, and non-Hispanic Whites. The significance of cataract-related visual impairment is emphasized along with its impact on daily activities, and the need for accurate projections and interventions. The 2004 Eye Diseases Prevalence Research Group study is cited as a foundational work that identified cataracts as a leading cause of vision loss in the US.

The study's methodology involves the development of a custom forecast model using linear regression. This model utilizes cataract prevalence data from 2014 to 2019 sourced from the Vision & Eye Health Surveillance System (VEHSS) at the CDC. The data was processed using the Pandas library in Python, and Scikit-learn's Linear Regression algorithm was employed for pattern recognition. The study's projections for cataract prevalence in 2030, 2040, and 2050 are presented, showcasing an upward trend across all racial groups, meaning that the prevalences in the coming years of cataracts will see a huge increase.

As shown in the Figures, the results of the forecast model indicate a substantial increase in cataract prevalence among Asians, non-Hispanic Blacks, Hispanics, and non-Hispanic Whites by 2030, 2040, and 2050. The projected prevalence rates illustrate the potential burden of cataracts on public health, with the rates varying across racial groups. In 2030, the Asian community will see about a 16% increase from 2019 to 2030, and in 2040, it will increase an additional 16%, followed by another increase of the same rate in 2050. Experiencing the greatest rate of increase from 2019 to 2050, the cataract prevalence in the non-Hispanic Black population is expected to increase by a total of 55% nationwide. Similarly, the cataract prevalence in the Hispanic population will increase nearly 38% from 2019 to 2050. Furthermore, the study highlights the varying rates of increase in cataract prevalence among different racial groups during different intervals.

The research paper offers a data-driven analysis that projects cataract prevalence trends for 2030, 2040, and 2050 among different racial groups in the US. By employing advanced statistical modeling techniques, the study highlights the potential challenges posed by cataract prevalence and underscores the significance of targeted interventions and public health strategies.

6. References

Richter, Grace M, et al. "Prevalence of Visually Significant Cataract and Factors Associated with Unmet Need for Cataract Surgery: Los Angeles Latino Eye Study." Ophthalmology, U.S. National Library of Medicine, Dec. 2009, www.ncbi.nlm.nih.gov/pmc/articles/PMC2787839/.

4"Vision and Eye Health Surveillance System (VEHSS)." Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, ddt-vehss.cdc.gov/. Accessed 17 Aug. 2023

The Eye Diseases Prevalence Research Group*. "Prevalence of Cataract and Pseudophakia/Aphakia among Adults in theUnited States." Archives of Ophthalmology, JAMA Network, 1 Apr. 2004, jamanet-work.com/journals/jamaophthalmology/fullarticle/416230

Shahbazi, Sara, et al. A Cross-Sectional Retrospective Analysis of the Racial and Geographic Variations in Cataract Surgery, PLOS ONE, journals.plos.org/plosone/article?id=10.1371/journal.pone.0142459# abstract0.

CDC. "Vision and Eye Health Surveillance System (VEHSS)." Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 2014, ddt-vehss.cdc.gov/LP?++Level1=Cataract&Level2=Cataract%2BPrevalence&Level3=Diagnosed%2BCataract%2C%2BAny%2BT++ype&Level4=Diagnosed%2BCataract-++Any%2BStage&LocationId=&DataSourceId=MEDICARE&IndicatorId=QDXDC6~R6_ALL&ShowF++ootnotes=true&View=NationalMap&CompareViewYear=1&CompareId=&CompareId2=&Yea++rId=YR11&ResponseId=R6_ALL&AgeId=AGEALL&GenderId=GALL&RaceId=ALLRACE&RiskFa++ctorId=RFALL&RiskFactorResponseId=RFTOT&DataValueTypeId=CRDPREV&MapClassifi++erId=quantile&MapClassifierCount=4&CountyFlag=N.