Assessing Underrepresentation in Machine Learning Datasets for Cardiovascular Disease (CVD) Research

by

CJ Dunn, Danilo Correia, and Liam Rathke

Worcester Polytechnic Institute

Advisor: Emmanuel Agu

May 3, 2022

Index

0. Figures and Tables, Acknowledgements (pg. 0)

1. Abstract (pg. 3)

2. Introduction (pg. 4)

    2.1 Background (pg. 4)

    2.2 Related Work (pg. 6)

2.3 Methodology (pg. 7)

    2.4 Results (pg. 7)

    2.5 Project (IQP) Goals (pg. 7)

3. Related Work (pg. 8)

4. Methodology (pg. 10)

    4.1 Dataset Review (pg. 12)

5. Results (pg. 14)

6. Discussion (pg. 28)

7. Conclusion (pg. 30)

Figures and Tables

5.1 - Racial and ethnic distributions of reviewed study datasets

5.2 - Cumulative racial and ethnic distribution of datasets, including US Census

5.3 - Cumulative racial and ethnic distribution of datasets, excluding US Census

5.4 - Percentage of Caucasians for each dataset

5.5 - Percentage of Hispanics for each dataset

5.6 - Percentage of Africans for each dataset

5.7 - Percentage of Asians for each dataset

5.8 - Percentage of subjects of other races and ethnicities for each dataset

5.9 - Differences in representation across racial and ethnic groups

5.10 - US population by racial and ethnic groups, according to Census data

5.11 - Racial and ethnic distribution in the NHANES dataset

5.12 - Racial and ethnic distribution in CDC Cardiovascular Disease Study 19-20

5.13 - Racial and ethnic distribution in the NHANES 17-18 dataset

5.14 - US national gender distribution, according to Census data

5.15 - NHANES dataset gender distribution

Acknowledgements

We would like to thank our Interactive Qualifying Project advisor, Professor Emannnuel O. Agu, who helped us throughout the project period and provided invaluable advice and feedback at every step.

1. Abstract

Cardiovascular diseases (CVDs) are the single most common global cause of death. With recent advancements in computer technology, machine learning (ML) and artificial intelligence (AI) are now common research aids for CVD research. In this paper, we investigate potential underlying biases in the data used by CVD studies that leverage ML and AI. After analyzing 11 CVD datasets, we found three datasets which included parameters for race/ethnicity and gender, all of which were demographically consistent with the US Census. However, the remaining 7 datasets referenced neither race/ethnicity nor gender. CVDs manifest differently across race/ethnic and gender groups, and thus research using datasets with unclear demographics could lead to inaccurate results. More investigation is necessary to quantify the impact of misrepresentation across demographic groups in CVD research.

2. Introduction

2.1 Background

Cardiovascular diseases (CVDs) include various symptoms, ailments, and illnesses related to heart function. While strokes and heart attacks are well-known CVDs, the disease class also includes many forms of heart disease, and conditions events such as atrial fibrillation. Put together, CVDs represent the largest cause of death globally; in 2012, an estimated 17.5 million people died as a direct result of CVD [5], rising to 17.7 million by 2017 [3], while the true number could be even larger. In the United States, 30% of deaths are attributed to CVD, and this number rises to 45% for the European Union [4].

Traditionally, CVDs are monitored and tracked individually by physicians [4] - thus, getting an early diagnosis is contingent on access to a physician with knowledge of CVD warning signs. While there are many risk factors that can predispose an individual to CVDs, studies show that over 70% of cases are directly linked to a subset of causes, such as hypertension, tobacco use, high non-HDL cholesterol, and a poor diet [3]. Other parameters can also reliably predict eventual CVD likeliness: better-educated people are less likely to die of CVDs, and the inverse relationship between education and CVD death rate is stronger than the correlation of wealth/income and CVD diagnosis [3]. In short, it is possible to identify the individuals most likely to succumb to CVD with reasonable accuracy; improving the accuracy of CVD prediction, or expanding access to CVD screenings could help larger population segments receive the treatment they need earlier, potentially reducing the substantial CVD death rate.

The data-driven nature of CVD diagnosis suggests a natural link between CVD prediction and a rapidly-expanding field in computer science: machine learning. Indeed, with a relatively low barrier to entry, and a large array of CVD-related datasets ready for analysis, the number of clinical journals leveraging machine learning algorithms has increased in recent years [6]. As a baseline metric, a search made on May 2, 2022 of the JSTOR research database found 626 articles published in health-focused journals relating to “cardiovascular disease” and “machine learning” between 2011 and 2016, a 53% increase over the 401 articles published between 2001 and 2006.

There are two main practical applications of machine learning in the CVD field. First, machine learning for CVD prediction: collecting datasets including CVD-specific risk factors for populations, and creating a model that predicts the individuals in the population most likely to suffer a CVD illness in the future. Many such models exist, and the dataset-driven nature allows for cross-verification across algorithms; for example, an algorithm built from one dataset could be tested against a different dataset, validating or invalidating said algorithm’s results, and potentially providing new insights for the second dataset. Machine learning researchers are actively comparing the performance of leading CVD algorithms, with the general consensus that machine learning algorithms are better at predicting CVD than other conventional methods [9]. Second, machine learning for CVD research: using machine learning techniques while researching cures for CVD illnesses. For example, cures for atrial fibrillation, a heart condition that leads to blood clots, are often unreliable, in part because identifying important electrical signals is difficult. Using machine learning, researchers created an algorithm to identify said electrical signals with over 95% accuracy [2]. In both cases, machine learning techniques offer a promising solution to CVD-related challenges. First, computer algorithms can produce more consistent results, meaning that good machine learning algorithms will be more precise when predicting CVD. Second, computerized algorithms can be run on large groups automatically, increasing accessibility for and awareness of CVD warning signs.

While the potential for machine learning in medicine, and more specifically, CVD, is great, there are many complications that must be taken into account. For instance, medical journals employ statisticians to review papers for accuracy before publication, but not all statistical editors are trained in machine learning. Without adequate support for this new technology, machine learning papers submitted to public health journals aren't subjected to a usual standard of review [6]. More importantly, there is evidence that CVD warning signs and results vary across racial and ethnic groups [1]. Thus, biased datasets used in machine learning studies could lead researchers to biased conclusions. Although the main focus of many CVD papers leveraging machine learning involves improving algorithms for CVD prediction, instead of applying machine learning directly to research, an algorithm optimized on biased or unrepresentative data could steer public health officials towards suboptimal CVD diagnosis and treatment strategies. In turn, minorities might receive inadequate diagnoses and care.

2.2 Related Work

Several papers related to CVD and ML identified race and ethnicity as a key factor when predicting cardiovascular disease in individuals. In 2015, researchers discovered that including race and ethnicity significantly improved cardiovascular prediction accuracy (*Race/Ethnic Differences in the Associations of the Framingham Risk Factors with Carotid IMT and Cardiovascular Events*) [1]. Two years later, another paper (*Can machine-learning improve cardiovascular risk prediction using routine clinical data?*) found that race and ethnicity was the best predictor of CVD in individuals when using the logistic regression model [5].

2.3 Methodology

In this Interactive Qualifying Project, we seek to understand whether the dataset-driven approach for machine learning research on cardiovascular disease is dependent on biased study results. First, we conducted a literature review of current materials related to CVD, machine learning, and CVD research using machine learning. Second, we collected 11 standard CVD datasets. Third, we contacted CVD and machine learning researchers to gain additional data and advice. Fourth, we created a data pipeline to run statistical analyses on the received CVD datasets.

2.4 Results

Of the 11 datasets collected, only three included race/ethnicity and gender parameters. These three datasets were demographically consistent with US Census results, with the exception of a dataset scaled to represent prevalent racial and ethnic groups in the US equally. The remaining seven datasets did not include race/ethnicity and gender parameters, suggesting that demographic imbalances might exist. These imbalances could lead to skewed algorithms, potentially leading to inequitable treatment for certain demographic groups.

2.5 The goal of the Interactive Qualifying Project (IQP)

- Determine the extent to which minority groups are underrepresented in cardiovascular disease research datasets
- Review CVD studies to understand underlying factors that might lead to differing representation across demographic groups

3. Related Work

Differences in the way cardiovascular diseases manifest themselves across racial and ethnic groups are already well-known. Specifically, a 2015 study (*Race/Ethnic Differences in the Associations of the Framingham Risk Factors with Carotid IMT and Cardiovascular Events*) created a model that predicted cardiovascular risk when considering race and ethnicity alongside other typical heart disease parameters [1]. This model was more accurate than other models that excluded race and solely based their predictions on risk factors. In other words, in order to deliver adequate preventative care for CVD ailments, a nuanced understanding of racial and ethnic background is critical. Similar to our project, researchers worked with datasets that included additional demographic information, finding racial and ethnic factors extremely useful when making a CVD prediction. However, this study focused on improving CVD prediction algorithms instead of assessing the state of diversity and representation in CVD datasets.

Surprisingly, numerous recent journal articles focused exclusively on CVD prediction with machine learning make little or no reference to race or ethnicity at all, in part a result of the common *Heart Disease Data Set*, which itself does not reference race or ethnicity. However, including race and ethnicity as ML dataset parameters can often yield stronger results. A 2017 journal article (*Can machine-learning improve cardiovascular risk prediction using routine clinical data?*) used a cohort of almost 300,000 to compare the effectiveness of four different machine learning algorithms in CVD prediction: logistic regression, random forest, gradient boosting machines, and neural networks. Ethnicity was the number one predictor for CVD in the logistic regression model, and the number three predictor for CVD with the random forest, gradient boosting machine, and neural network algorithms [5]. Thus, within machine learning studies, racial and ethnic background information is a key component to a good CVD prediction. Like the previous paper, this study identified the importance of accounting for race and ethnicity under the context of improving prediction algorithms, and didn't conduct a comprehensive review of the state of diversity in cardiovascular disease research using ML.

Much of our research consisted of identifying datasets used within CVD studies; the most relevant datasets are identified and further discussed in section 4.1 (dataset review).

4. Methodology

This section presents the tasks that were completed in order to accomplish the goal of the research in addition to the methods appropriate to achieve each goal. Due to the complexity of factors pertaining to the topic, a multifaceted approach was taken in order to understand the scope of the project. This approach included: 1) Gathering background information by accumulating research papers of value within the topic of machine learning and artificial intelligence concerning cardiovascular disease. 2) Collecting the datasets used within the accumulated research papers. 3) Email correspondence with authors of accumulated research papers. 4) Developing a data pipeline in order to further break down data. These steps were necessary to develop a complex understanding of the variables affecting our research.

Accumulating research papers within the field allowed for a basic understanding of goals within the topic. By understanding the research of others, we were then able to ascertain the underlying deficiencies within this field of study. Collecting datasets from valued papers allowed for a data pipeline to then be crafted. With the creation of said pipeline, comparisons between research value and validity were able to be made. Email correspondence with authors allowed for the collection of author opinions on the weak and strong points of their own research as well as potential hidden issues with the underlying structure of their papers. This structure for gathering information from a plethora of sources allowed for both an analytical objective approach to understanding the story of the data as well as a potentially biased version coming from authors. However, these methods of gathering information had varying levels of success.

Gathering data began by accumulating studies focused on machine learning as a viable way to further diagnose, perceive, and treat a plethora of different cardiovascular diseases. During the accumulation of datasets, study frameworks such as clinical trials, experimental studies, longitudinal studies, and case studies were all avoided. These frameworks focus on variable manipulation, time variations, or individual progression; whereas the goal of our data gathering was to find the distribution of gender and race within a dataset. Otherwise, implying that each individual within the dataset was treated the same would be impossible. For this, we found correlational studies to be the best fit. Additionally, many studies use secondary datasets acquired through national health databases rather than using primary data acquired themselves.

During the collection of datasets from machine learning research projects pertaining to cardiovascular disease there were no ethical considerations involved in our choices. The papers we chose datasets from needed to meet a minimum requirement of being peer reviewed to ensure validity. Additionally, while the reliability of secondary datasets were confirmed by the peer review process as well, truly authenticating medical datasets is troublesome. There are common human errors (encoding) and also negligence (duplicates). Through the process of searching for the best possible dataset, meaning the most commonly used because of its ability to meet authentication standards, we found that the Cleveland dataset as well as the SwitzDataset to be the gold standard for machine learning studies relating to cardiovascular disease. Despite these qualifications, neither dataset include the attribute of race within their scope [14].

After acquiring a diverse set of datasets from various sources, we began crafting a data pipeline for analysis. Before analysis, datapoints whose origins were shared were grouped into broader terms. These grouping terms were decided upon after the data points race origins could be traced. For example, “African Americans” were grouped into “of African descent” which was then joined into “Black” after finding this is the most common dataset attribute descriptor. The framework of the data pipeline is a simple comparison breaking down the proportions of race and gender against the proportions of census data for the area of coverage implied by the sample. For samples of the national scale, national proportions were used, while for samples on a local scale, local proportions were used. This comparison of experimental versus actual was then further broken down into a range of charts and graphs indicating the representations of our targeted attributes within the data.

Statistical analysis was then done in order to indicate trends for every accounted for attribute. This statistical summary was done for each individual dataset as well as for each attribute by accumulating the range of errors. The purpose of this multifaceted approach was to not only see trends within an individual study, but be able to see trends on a broader scale for gender and race within the field of study.

4.1 Data

NHANES 2011-2016:

This dataset was not originally intended for machine learning. Data was accumulated for medical research purposes, and only later adopted by projects attempting to model CVD in a predictive manner. Although there are dozens of attributes tracked in the dataset, no studies attempting to make predictive models used more than 11 attributes. The most common attributes included cardiovascular disease indicator attributes and identifying attributes such as age, gender, and race.

NHANES 2018:

This dataset was also not originally intended for machine learning. This dataset was created with the same intentions and specifications as NHANES 2011-2016 and was similarly adopted later as a means to make predictive models for CVD. Once again the attributes examined for the purpose of CVD models were CVD indicator variables and identifying attributes such as age, gender, and race.

CDC:

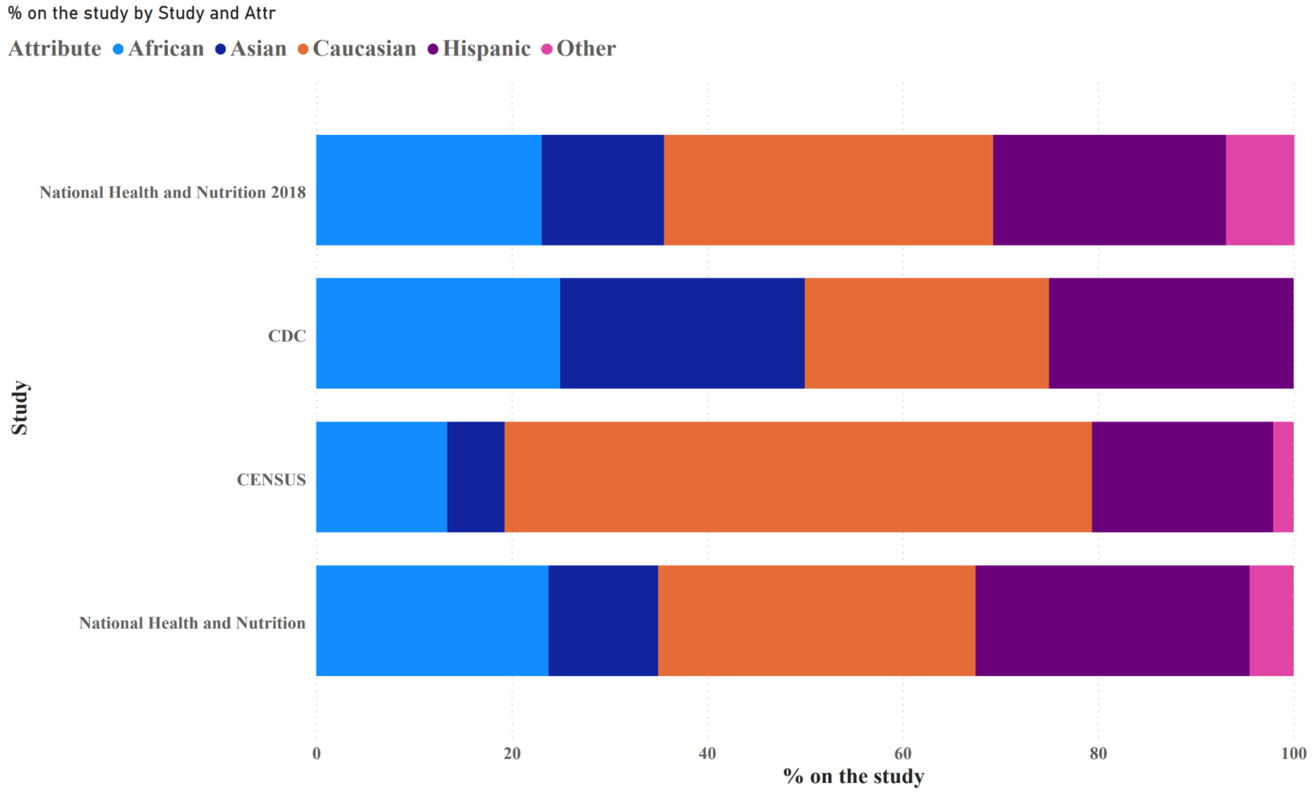
This dataset was purposefully crafted from a larger grouping of surveys to be used for artificial intelligence training and modeling. For this reason, the dataset shows an equal distribution of demographics. As a result of the focus on enabling i=i comparisons with US census data, minority groups within this study are massively overrepresented. For the purpose of predictive modeling, this data equality allows for the attribute of origins to be a key factor for accurate modeling.

Heart and Disease Data Set:

This dataset is the self-proclaimed most commonly cited data source within studies using machine learning and artificial intelligence to produce models pertaining to cardiovascular disease. This dataset is not included within this paper's data modeling as the attributes of gender and race are not accounted for within this set. This dataset is an accumulation of multiple other data sources spanning the globe [14].

5. Results

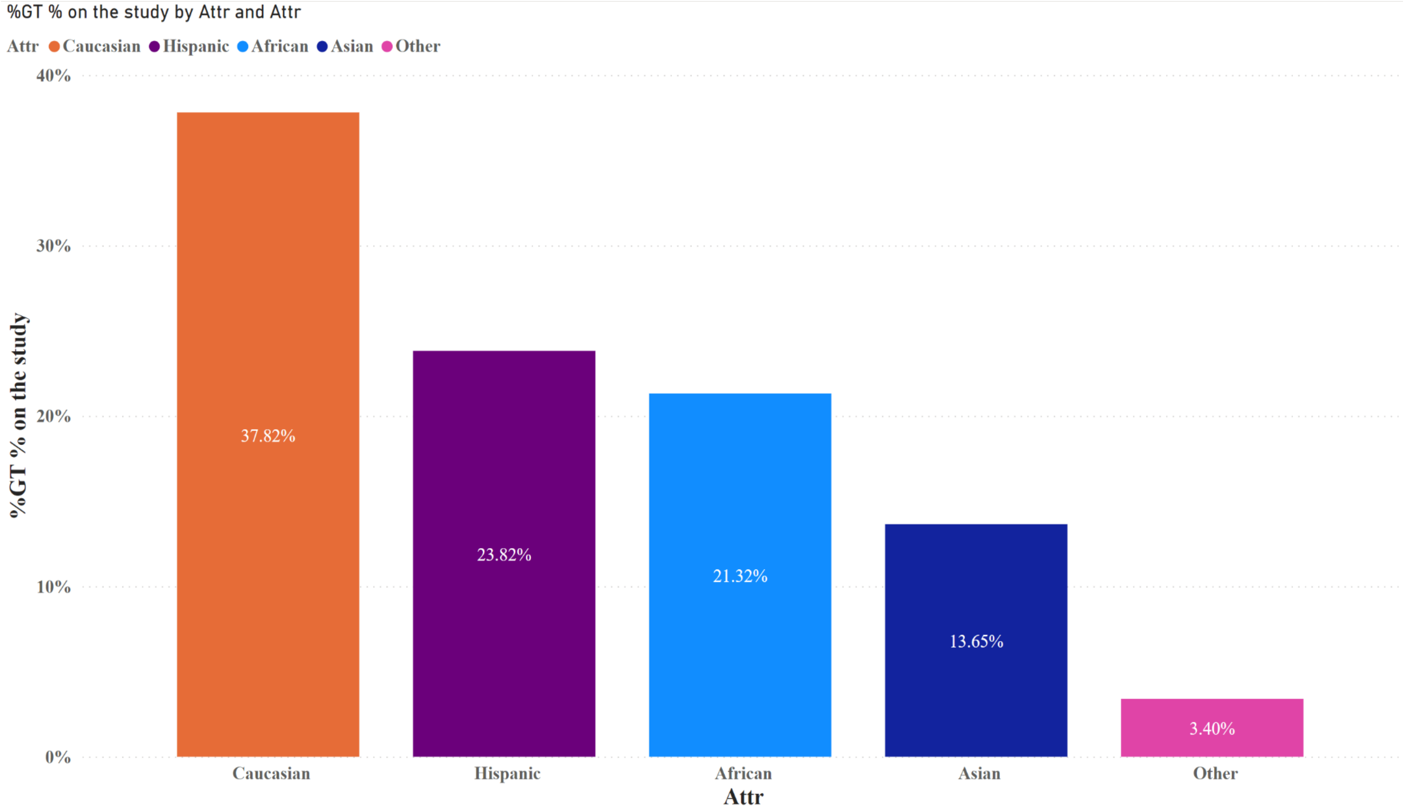
For better analysis and comparison, we decided to compare the datasets using Power BI technology to create data visualizations. The goal of the visualizations is to facilitate human perception regarding the difference in proportions for each ethnicity for each dataset, thus relating these differences with society's issues and structure. The following figures were generated using Power BI data interpretation and report generator, along with some custom queries and data input.



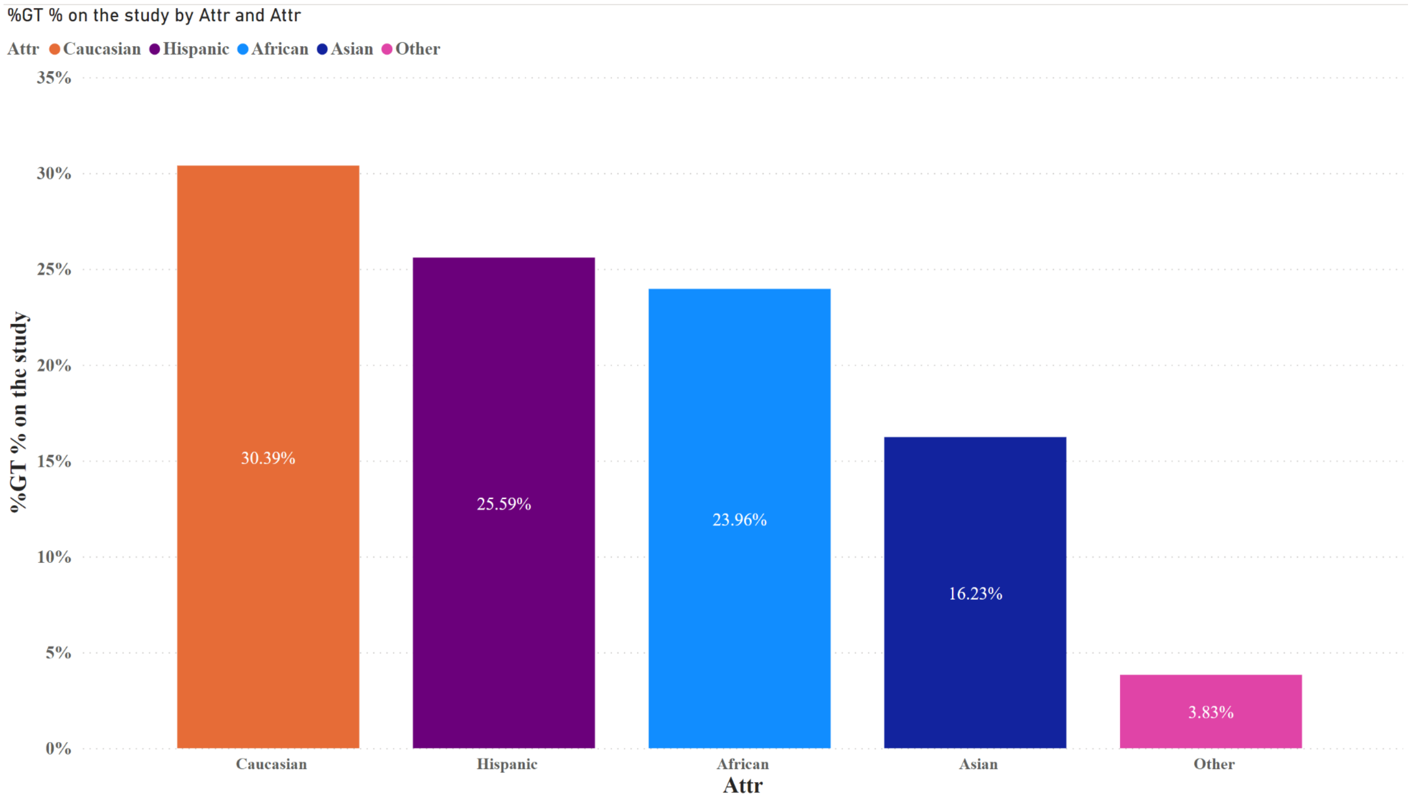
(figure 5.1 - percentage of each race for each study dataset - CENSUS for control)

In order to identify how studies are statistically biased regarding their datasets, we used the American CENSUS as a control study that represents the real ethnicity distribution in the US population. Notice that, from figure 5.1, Caucasian and Hispanic populations are the ones that are mostly underrepresented when compared to the general population distribution from the CENSUS. The African population, on the other hand, is overrepresented in percentage wise. There are social, economic, and technical reasons why the datasets are following this pattern of over-representation of the usually underrepresented population that we are going to hypothesize later on in this paper.

The following two figures project the distribution of the ethnicities for all studies, with and without adding the CENSUS to it.



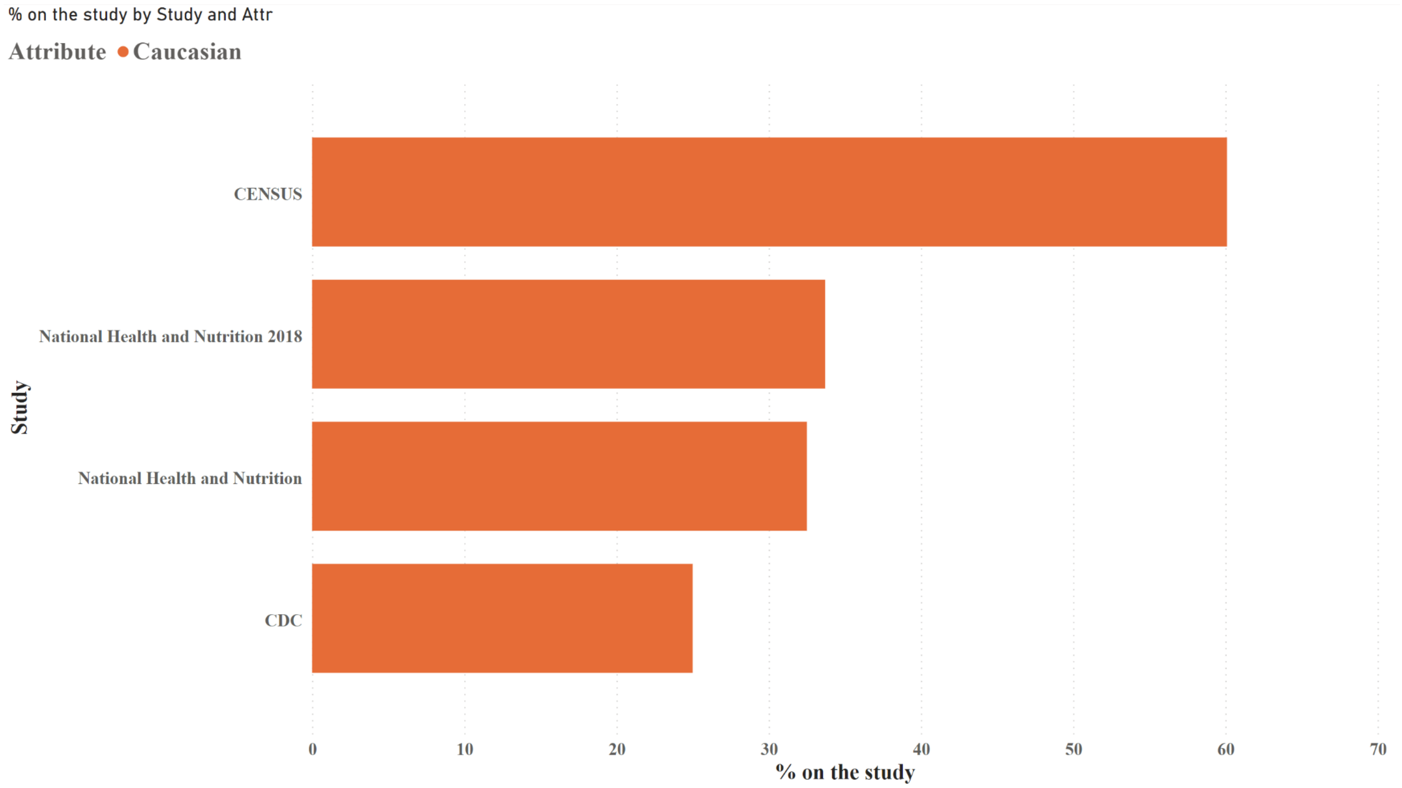
(figure 5.2 - percentage of each race in each study, including the CENSUS)



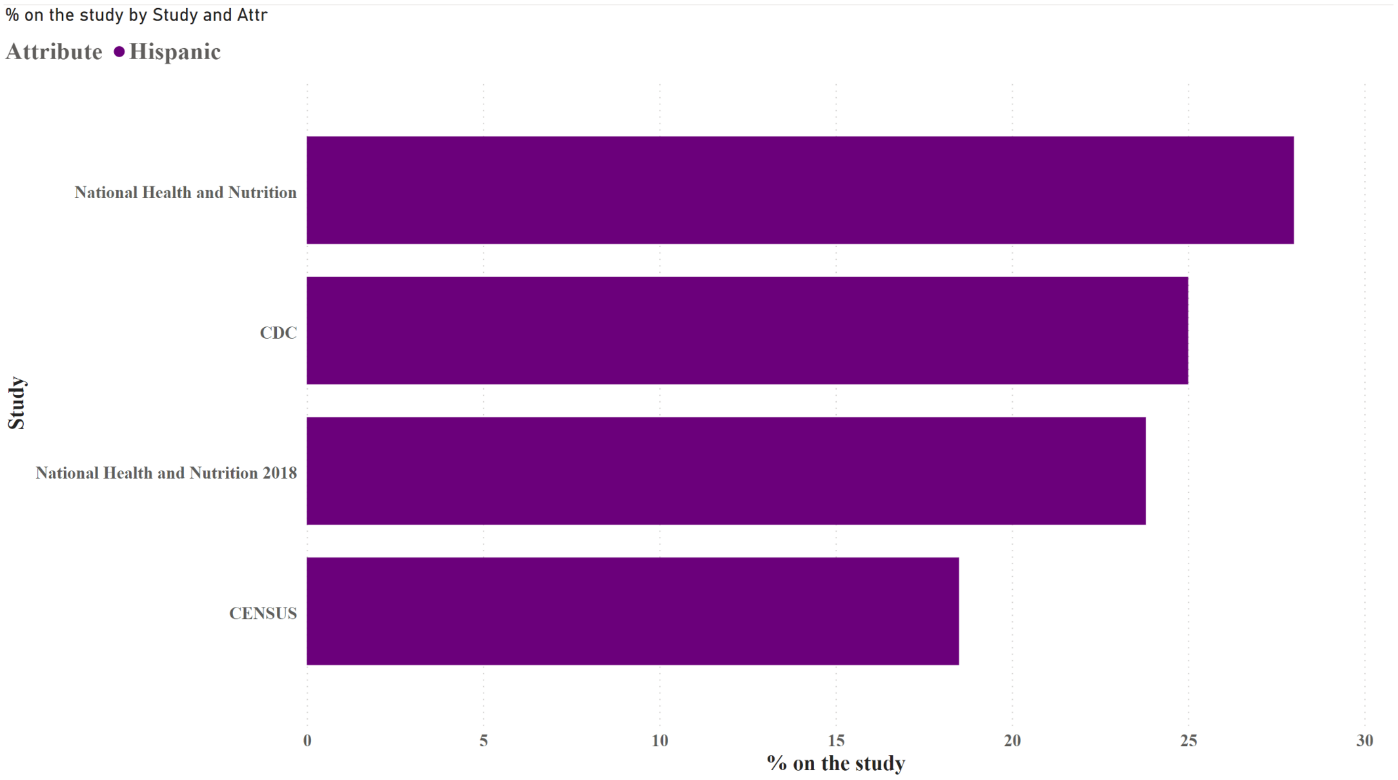
(figure 5.3 - percentage of each race in all studies - no CENSUS included)

Closely analyzing each population, we can notice that Caucasians constitute more than 35% (figure 5.2), on average, of the sum of populations of all datasets. When the census is taken out of the analysis, Caucasians are now 30.39% of the total population while other races slightly increase (figure 5.3). There is also an almost constant Asian population percentage independently if the census is being considered or not.

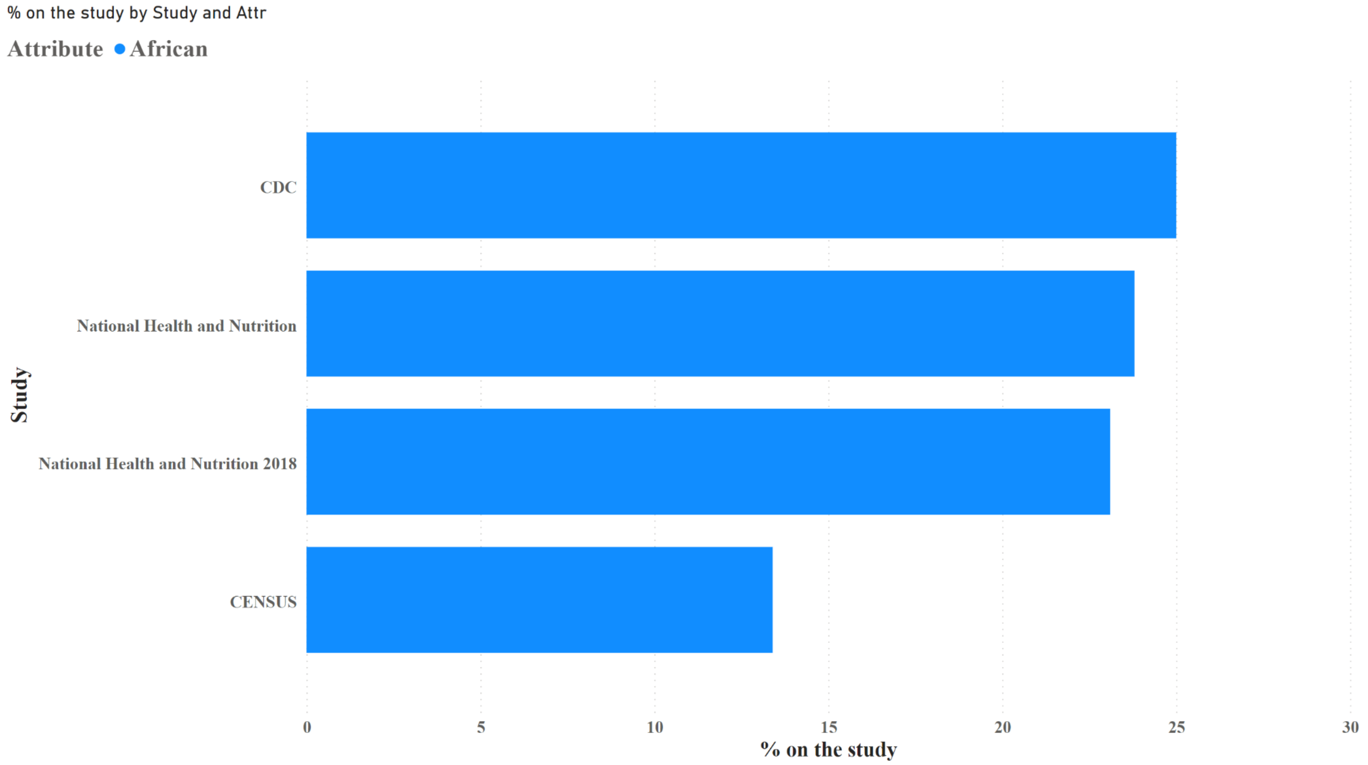
For a better analysis, we can also break down the percentages for each race for each study:



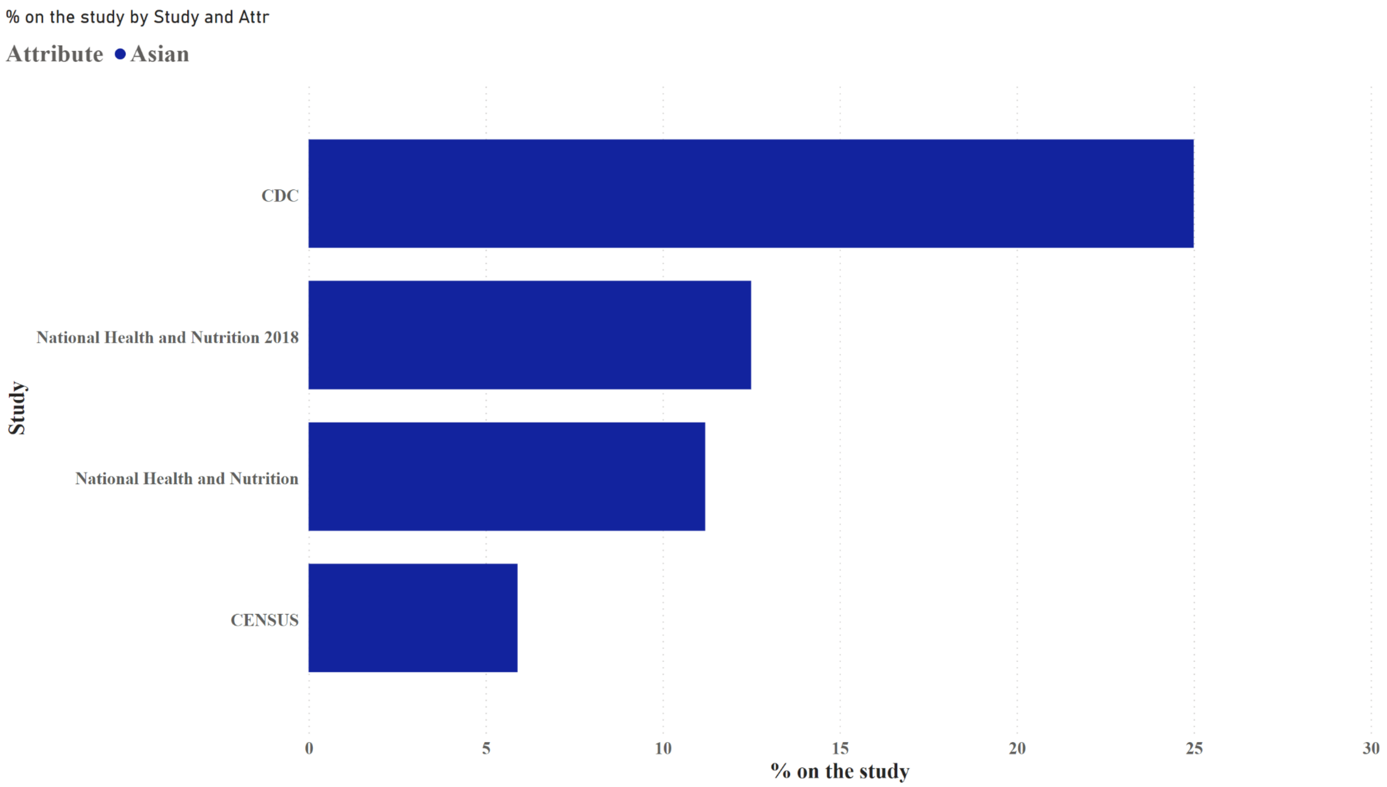
(figure 5.4 - percentage of Caucasians for each dataset)



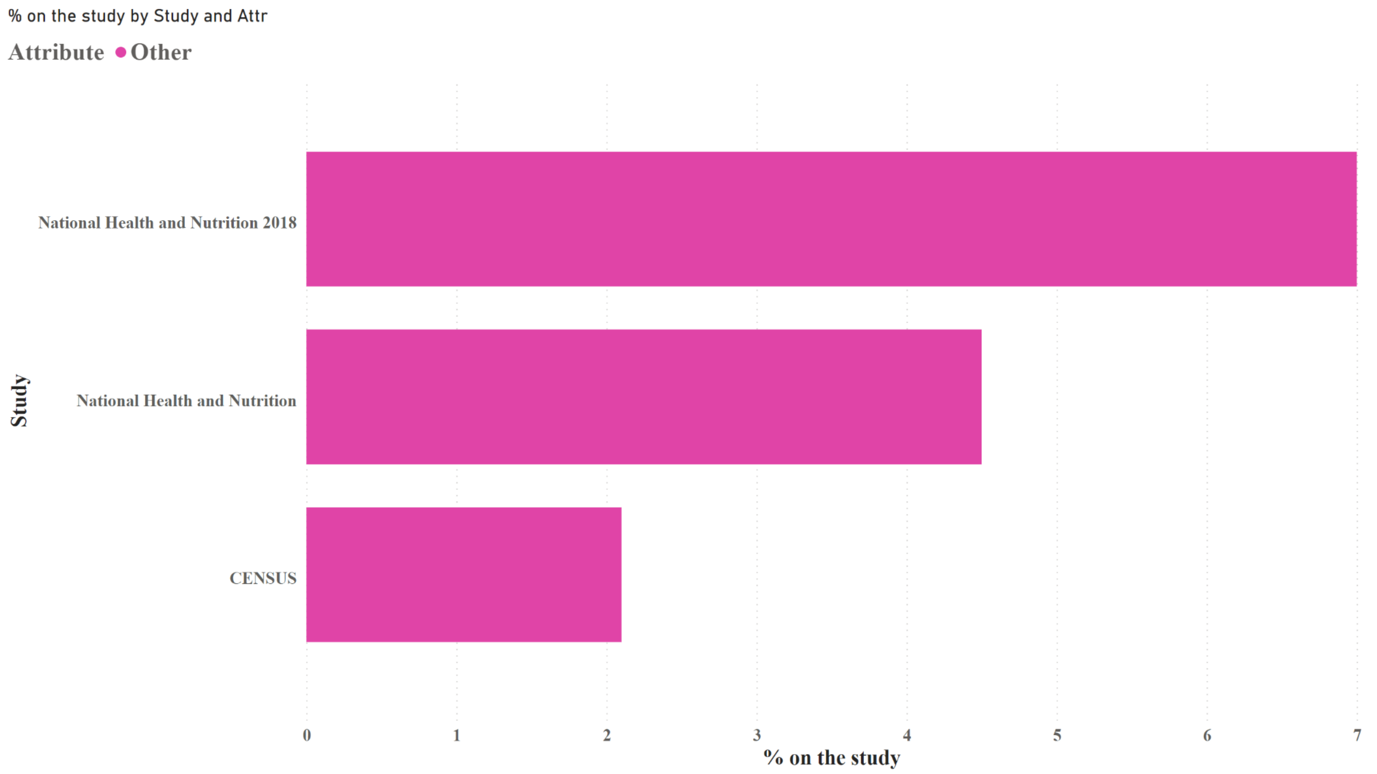
(figure 5.5 - percentage of Hispanics for each dataset)



(figure 5.6 - percentage of Africans for each dataset)



(figure 5.7 - percentage of Asians for each dataset)



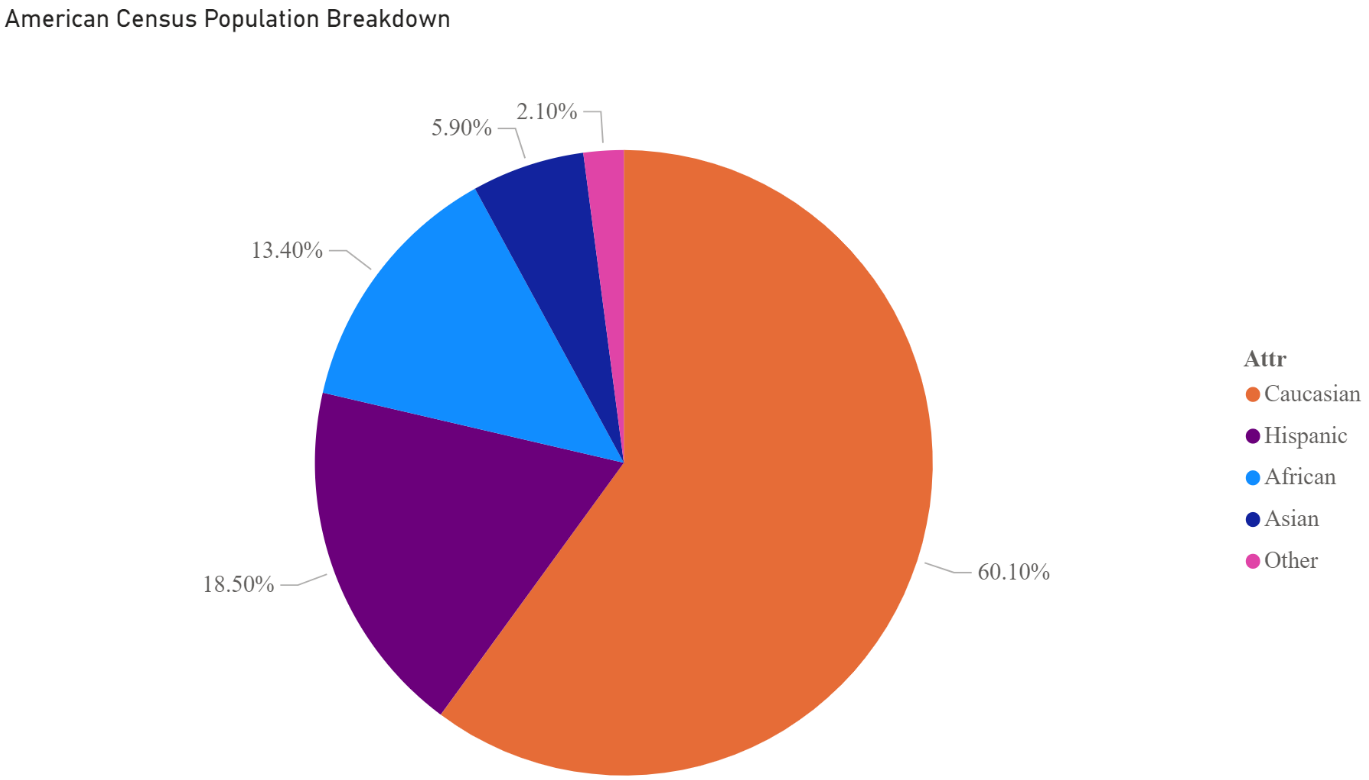
(figure 5.8 - percentage other races for each dataset)

After breaking down the different races for each dataset population, the gap between datasets and the real population is more visible. For each race, we decided to create the following table comparing the values of the average percentile used in the studies and the percentile in the US population from the census:

Race	**Average percentile used in the studies datasets**	**Percentile from true population (census)**	**Difference**
Caucasian	30.4	60.1	-29.7
Hispanic	25.6	18.50	7.1
African	23.9	13.4	10.56
Asian	16.23	5.9	10.3
Others	5.75	2.1	3.65

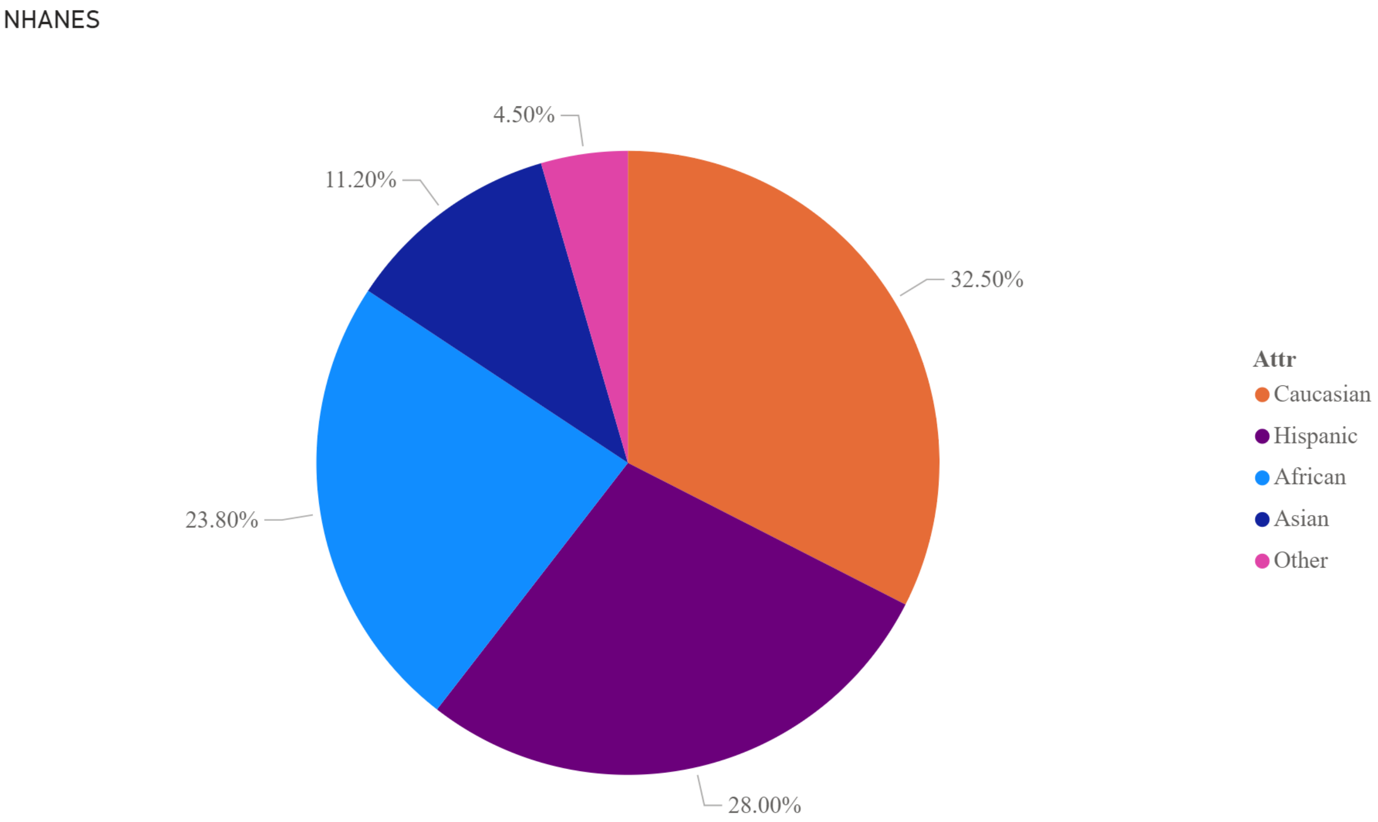
(Table 5.9 - differences in representation of different races)

In conclusion, there is a tendency to equally distribute the different races for the datasets used for the machine learning models that are mostly used for these types of prediction. This "flattening" is directly correlated with how the algorithm is trained. This will be further discussed in the discussion section.



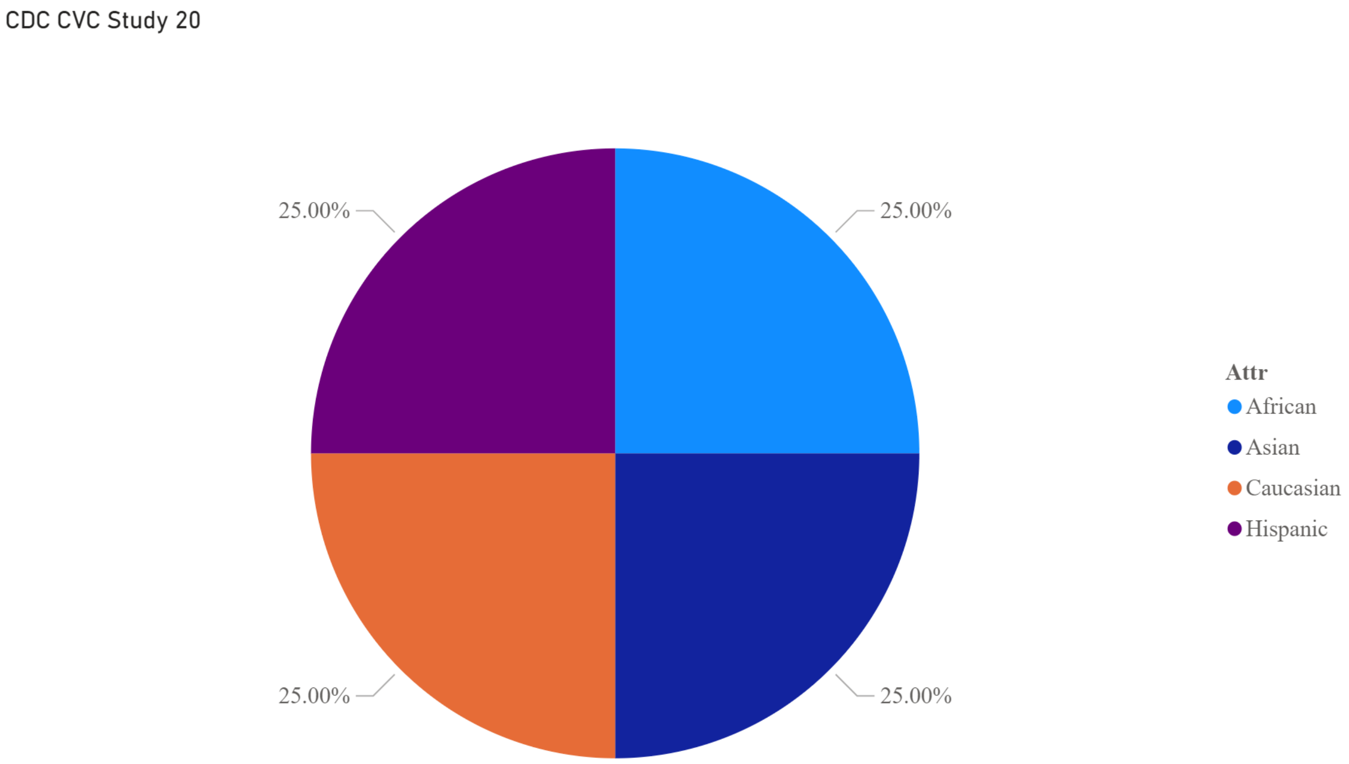
(figure 5.10 - American CENSUS for control)

This pie-chart is a population breakdown in the United States by race. In this graph "Other" represents a plethora of origins including but not limited to: Pacific Islander, Native American, and multicultural people who identify as such. While these percentages can be represented within this graph the same can not be said for the graphing of cardiovascular disease datasets which lack these options. For this to be a true control graph these origins have been labeled "others"s.



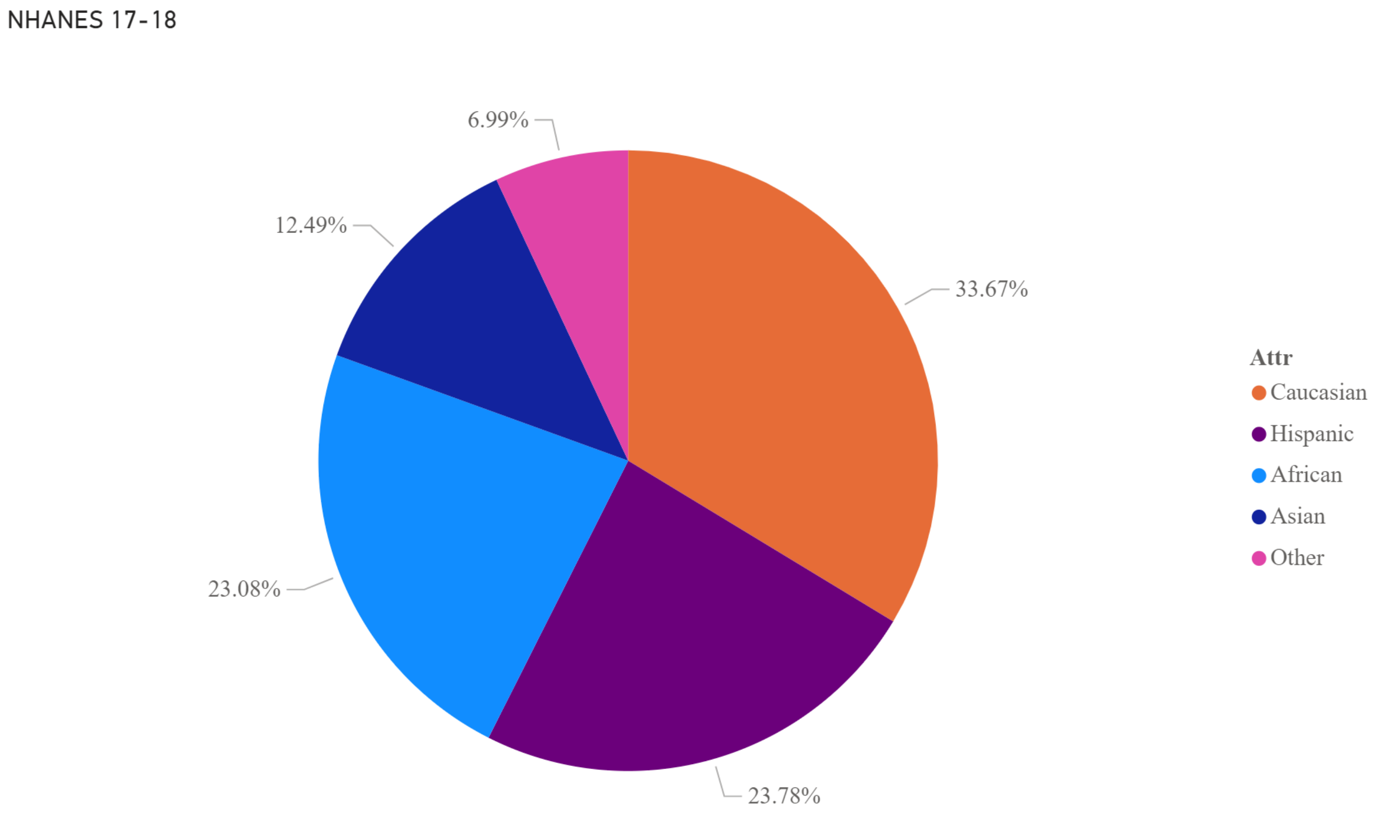
(figure 5.11 - NHANES Study 2011-2016: Race)

This pie-chart is a representation of the races described within the NHANES (National Health and Nutrition Examination Survey) dataset taken from 2011-2016. This dataset when compared with the census data for the same time-period shows an over-representation of all minority groups by a maximum of 216.5%, a minimum of 150%, an average of 83.8%



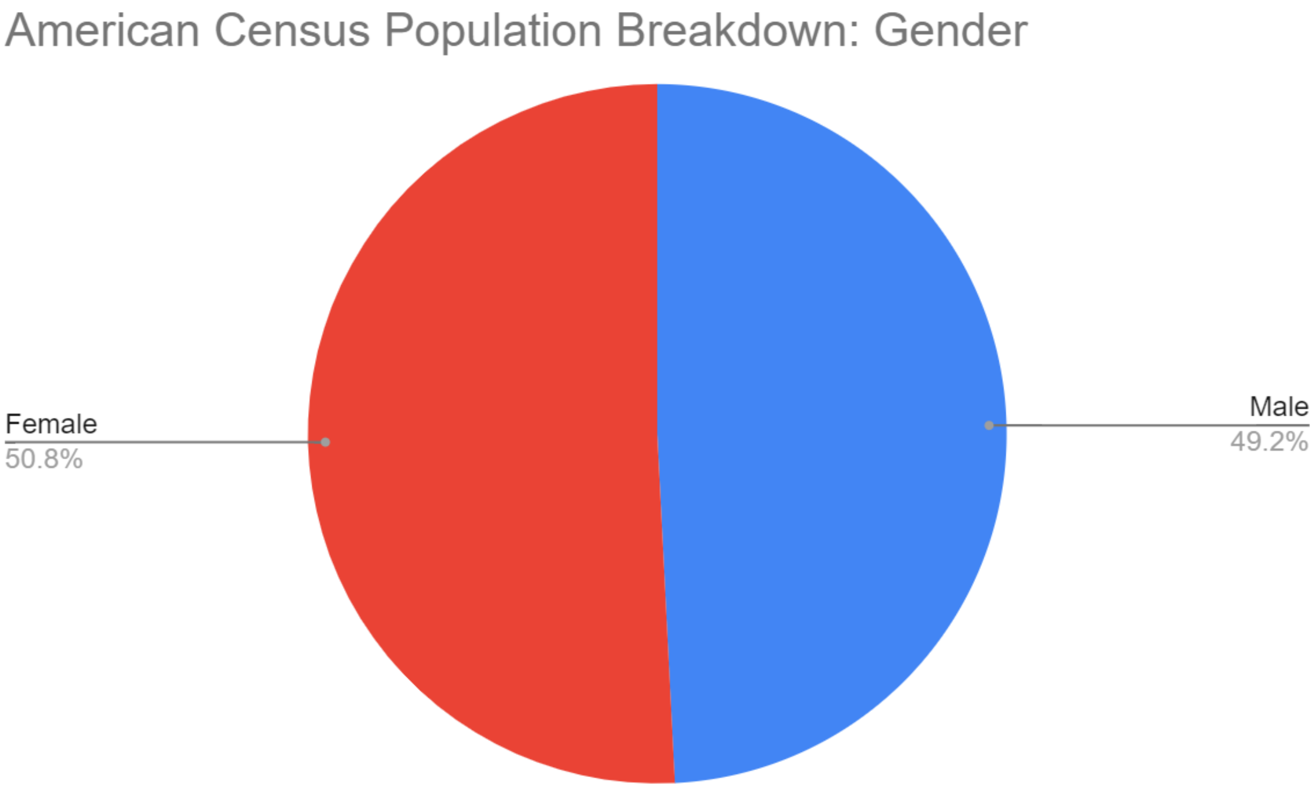
(figure 5.12 - CDC Cardiovascular Disease Study 19-20)

This study depicts, much like figure 5.3, an overrepresentation of minorities within the framework of an American population study. This sample was stated to be “purposely equal” without further explanation. The implications of this statement are that this dataset was purposefully crafted with the intention of using it for a type of artificial intelligence model training called classification.



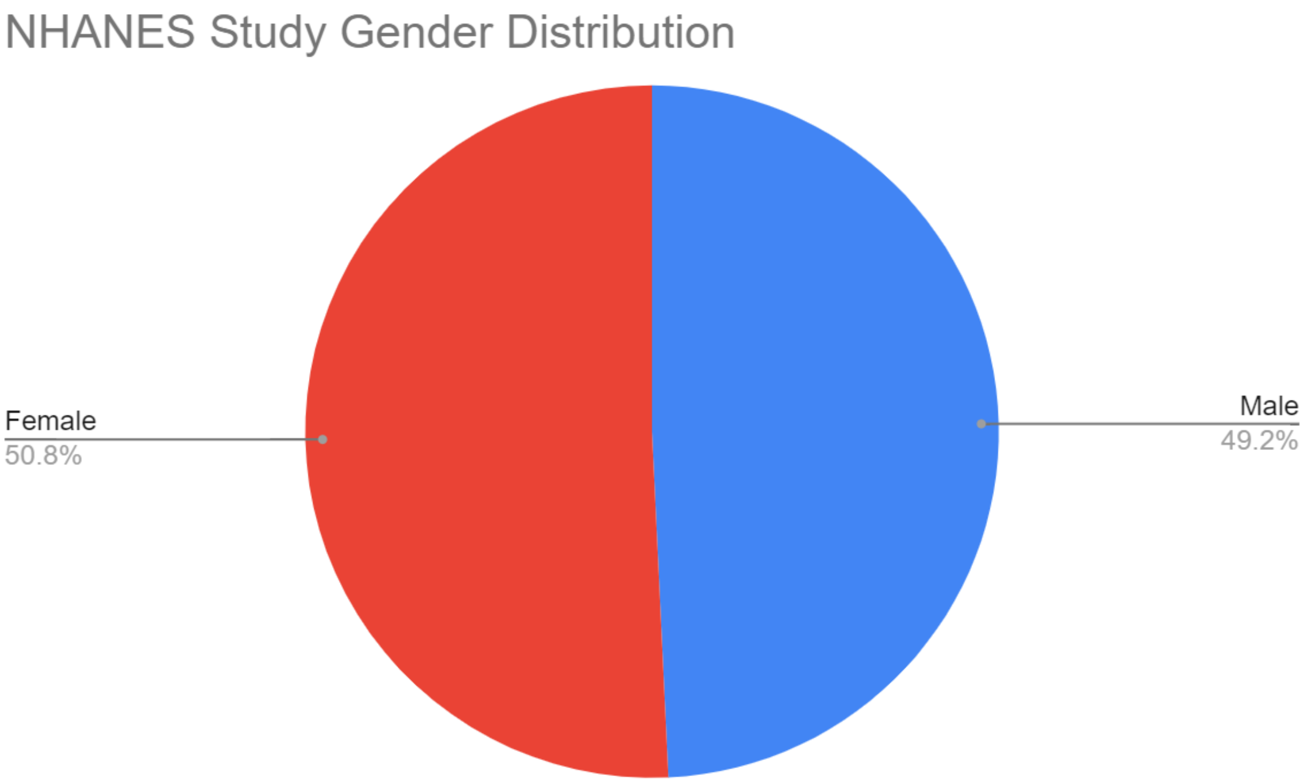
(figure 5.13 - NHANES Study CVD 17-18)

This sample when compared with census data of the same year shows an average over-representation of minority groups at a rate of 212%. This value is skewed by the outlying 330% over representation within the “Other” attribute. This over-representation allows for easier classification within the machine learning models. These models being trained with a “flat” or even representation are stronger than models trained with census consistent data.



(figure 5.14 - American Census Population Breakdown: Gender: 2016)

Figure 5.6 is the gender distribution of America from the 2016 census. This is our control for distribution of samples taken on a national scale during this year.



(figure 5.15 - NHANES Study Gender Distribution: 2016)

Figure 5.7 is the gender distribution of the NHANES 2016 dataset. The distribution is the exact same as the distribution from the American census of this year. This consistency is valued within the sample when looking to make a predictive model for the other attributes within the dataset. The argument can be made that while this sample represents the national population as a whole, it fails to represent the sub-population affected by cardiovascular diseases.

Through investigating the most commonly used heart disease datasets found in machine learning studies the most glaring discovery found was that race and gender go mostly unaccounted for within these datasets and in perpetuity the models themselves. The UCI machine learning repository stores the *Heart Disease Data Set* which is a consolidation of the four datasets Cleveland, Hungary, Switzerland and VA Long Beach. This consolidation of four databases is the most commonly cited datasource found in machine learning pertaining to CVD [14]. It accounts for 75 attributes, but the number of attributes used within machine learning models is a subset of 14 attributes. None of these seventy-five nor the fourteen attributes are race or gender.

6. **Discussion**

Our research made clear the issues involved with demographics within this field. Researchers looking to use medical data to create accurate models are seemingly struggling to use data that meets diversity goals. We found three open source datasets that were able to accurately sample cardiovascular disease markers as well as demographic markers. The graphical breakdown supplied in the results section shows datasets well suited for machine learning models within this field. The populations were equally distributed facilitating interpretation of data regardless of the statistical strategy used by the machine learning algorithm [18]. These datasets have reinforced that sampling for medical markers while also sampling for demographics is viable. The overrepresentation of minorities is ideal for classification studies of this manner. Yet, a majority of the research being published within this field is still using older instances of data that does not meet the same diversity standards, and avoids mentioning demographics at all.

The fifty-seven machine learning models trained upon the *Heart Disease Data Set* account for a self proclaimed “majority of all machine learning endeavors" in this field [14]. Per an accumulation of studies dating back to 2011 describing “profound race-associated disparities among those who are affected by and die from cardiovascular disease”, not publishing demographics within a dataset in the medical field is unexpected [15]. With the general consensus of the medical field being that race and gender play massive roles in health and in perpetuity the health care received by an individual, a dataset that does not disclose its demographics is likely doing so because of its inability to reach diversity goals within the sample. Email correspondence was attempted with both the publishing author of the dataset and the two studies citing this datasource most recently, both in the year 2022. After two weeks of waiting, zero responses were received from the studies citing the dataset. Likewise, there was no response from the publishing author, but within the README the questions posed were answered. According to the author "the data was formed with the intentions for machine learning modeling finding correlation between cardiovascular disease markers"'. There was no mention of race nor gender, yet the author is stating that the data was accumulated for a very specific purpose. One can surmise from this that studies citing this datasource with different intentions are doing so knowingly.

The *Heart Disease Data Set* was published in 2014 after sampling from the years 2011-2013. Yet, today it is still the standard dataset used for modeling without the attributes of gender or race. This could possibly indicate that current research struggles to deal with these attributes as factors within modeling; hence why current models choose to shy away from current data with demographic breakdowns.

7. Conclusion

Underrepresentation continues to be a large and long-standing issue in society. In this paper, we set out to assess the impact of demographic bias in datasets used for cardiovascular research studies. We conducted a survey of recently published works and gained access to industry-standard datasets to compare the diversity in data used for cardiovascular disease research with national averages. Although three of the 11 datasets analyzed as part of the project included demographic information, the remaining seven did not. The prevalence of datasets without demographic breakdowns suggest a worrying trend - that current health research could come to potentially inaccurate conclusions based on biased data.

The impact of demographic bias for cardiovascular disease research specifically is especially significant; it is well-established that many CVDs impact demographic groups differently. Thus, studies which use datasets with unclear or biased representation could come to inaccurate conclusions, which might exclude entire population segments from receiving adequate care. While more research is needed to quantify the exact impact of underrepresentation or misrepresentation in recent CVD studies, it is clear that dataset transparency is necessary to ensure equitable access to and benefits from cardiovascular research leveraging machine learning.

7. Bibliography

[1]C. M. Gijsberts *et al.*, “Race/Ethnic Differences in the Associations of the Framingham Risk Factors with Carotid IMT and Cardiovascular Events,” *PLOS ONE*, vol. 10, no. 7, p. e0132321, Jul. 2015, doi: 10.1371/journal.pone.0132321.

[2]M. I. Alhusseini *et al.*, “Machine Learning to Classify Intracardiac Electrical Patterns During Atrial Fibrillation,” *Circulation: Arrhythmia and Electrophysiology*, vol. 13, no. 8, Aug. 2020, doi: 10.1161/circep.119.008160.

[3]S. Yusuf *et al.*, “Modifiable risk factors, cardiovascular disease, and mortality in 155 722 individuals from 21 high-income, middle-income, and low-income countries (PURE): a prospective cohort study,” *The Lancet*, vol. 395, no. 10226, pp. 795–808, Mar. 2020, doi: 10.1016/s0140-6736(19)32008-2.

[4]P. Bizopoulos and D. Koutsouris, “Deep Learning in Cardiology,” *IEEE Reviews in Biomedical Engineering*, vol. 12, pp. 168–193, 2019, doi: 10.1109/rbme.2018.2885714.

[5]S. F. Weng, J. Reps, J. Kai, J. M. Garibaldi, and N. Qureshi, “Can machine-learning improve cardiovascular risk prediction using routine clinical data?,” *PLOS ONE*, vol. 12, no. 4, p. e0174944, Apr. 2017, doi: 10.1371/journal.pone.0174944.

[6]K. Singh, A. L. Beam, and B. K. Nallamothu, “Machine Learning in Clinical Journals,” *Circulation: Cardiovascular Quality and Outcomes*, vol. 13, no. 10, Oct. 2020, doi: 10.1161/circoutcomes.120.007491.

[7]J. Coe and M. Atay, “Evaluating Impact of Race in Facial Recognition across Machine Learning and Deep Learning Algorithms,” *Computers*, vol. 10, no. 9, p. 113, Sep. 2021, doi: 10.3390/computers10090113.

[8]A. J. Larrazabal, N. Nieto, V. Peterson, D. H. Milone, and E. Ferrante, “Gender imbalance in

medical imaging datasets produces biased classifiers for computer-aided diagnosis," *Proceedings of the National Academy of Sciences*, vol. 117, no. 23, pp. 12592–12594, Jun. 2020, doi: 10.1073/pnas.1919012117.

[9]C. Tesche and V. Brandt, "Calling for a New Framingham," *JACC: Cardiovascular Imaging*, vol. 14, no. 3, pp. 626–628, Mar. 2021, doi: 10.1016/j.jcmg.2020.12.027.

[10]P. Li, Y. Hu, and Z.-P. Liu, "Prediction of cardiovascular diseases by integrating multi-modal features with machine learning methods," *Biomedical Signal Processing and Control*, vol. 66, p. 102474, Apr. 2021, doi: 10.1016/j.bspc.2021.102474.

[11]G. Parthiban and S. K. Srivatsa, "Applying Machine Learning Methods in Diagnosing Heart Disease for Diabetic Patients," *International Journal of Applied Information Systems*, vol. 3, no. 7, pp. 25–30, Aug. 2012, doi: 10.5120/ijais12-450593.

[12]L. Yahaya, N. David Oye, and E. Joshua Garba, "A Comprehensive Review on Heart Disease Prediction Using Data Mining and Machine Learning Techniques," *American Journal of Artificial Intelligence*, vol. 4, no. 1, p. 20, 2020, doi: 10.11648/j.ajai.20200401.12.

[13]M. D. McCradden, S. Joshi, M. Mazwi, and J. A. Anderson, "Ethical limitations of algorithmic fairness solutions in health care machine learning," *The Lancet Digital Health*, vol. 2, no. 5, pp. e221–e223, May 2020, doi: 10.1016/s2589-7500(20)30065-0.

[14]A. Janosi, W. Steinbrunn, M. Pfisterer, and R. Detrano, "UCI Machine Learning Repository: Heart Disease Data Set," *Uci.edu*, 2019. https://archive.ics.uci.edu/ml/datasets/Heart+Disease

[15]L. C. Brewer and L. A. Cooper, "Race, Discrimination, and Cardiovascular Disease," *AMA Journal of Ethics*, vol. 16, no. 6, pp. 455–460, May 2013, doi: 10.1001/virtualmentor.2014.16.6.stas2-1406..

[16]United States Census Bureau, "U.S. Census Bureau QuickFacts: United States,"

www.census.gov, 2020. https://www.census.gov/quickfacts/fact/table/US/PST045221

[17]CDC, “National Health and Nutrition Examination Survey,” *wwwn.cdc.gov*, Feb. 2020. https://wwwn.cdc.gov/Nchs/Nhanes/2017-2018/DEMO_J.htm

[18]I.-K. Yeo, “A new family of power transformations to improve normality or symmetry,” *Biometrika*, vol. 87, no. 4, pp. 954–959, 2000.

.