Title

Data Analytics of Electronic Medical Record to Study Racial Diversities in Cardiovascular Diagnosis and Treatment.

Authors

Panahiazar, Maryam
Chern, Yorick
Riojas, Ramon
et al.

Publication Date

2022-05-01

DOI

10.3233/shti220519

Peer reviewed

HHS Public Access

Author manuscript

Stud Health Technol Inform. Author manuscript; available in PMC 2023 July 13.

Published in final edited form as:

Stud Health Technol Inform. 2022 May 25; 294: 550–554. doi:10.3233/SHTI220519.

Data Analytics of Electronic Medical Record to Study Racial Diversities in Cardiovascular Diagnosis and Treatment

Maryam PANAHIAZAR[a,b,1], **Yorick CHERN**[c], **Ramon RIOJAS**[a], **Omar S. LATIF**[d], **Dexter HADLEY**[e], **Ramin E. BEYGUI**[a]

[a]Department of Surgery, Division of Cardiothoracic Surgery, School of Medicine, University of California San Francisco, San Francisco, CA, USA

[b]Bakar Computational Health Sciences Institute, School of Medicine, University of California San Francisco, San Francisco, CA, USA

[c]Department of Bioengineering, University of California Berkeley, Berkeley, CA, USA

[d]Department of Internal Medicine, Mayo Clinic, Rochester, MN, USA

[e]Department of Clinical Sciences, College of Medicine, University of Central Florida, Orlando, FL, USA

Abstract

The study of precision medicine that measures the effects of social, cultural, and environmental influences on health is essential to improve health outcomes. Race is a social concept used historically to divide, track, control populations, and reinforce social hierarchies. Beyond genetics, race is also a surrogate for other socioeconomic factors affecting patient outcomes. Our data analytics study aims to analyze the Electronic Medical Record (EMR) to study patients of different races in diagnosing and treating Coronary Artery Disease (CAD). We found no race discrepancies at the University of California San Francisco Medical Centers. This study opens several new hypotheses for further research in this crucial field.

Keywords

Electronic Medical Record; Racial Study; Cardiovascular Disease; Data Analytics; Precision Medicine

1. Introduction

The risk of developing Coronary Artery Disease (CAD) is markedly different in patients of various races. Precision medicine research promises several approaches to introducing human genomics at the root of racial health disparities. Sometimes, genomic approaches are insufficient to address the various social experiences that often correlate with social behavior. Discrimination in housing and employment, inadequate access to health insurance,

[1]Corresponding Author, Maryam PANAHIAZAR; Maryam.panahiazar@ucsf.edu.

and implicit and explicit biases in medical care, for example, substantially impact health outcomes, long-term health, and health disparities. Therefore, precision medicine studies that measure the effects of cultural, social, and environmental influences on health are essential to improve health outcomes. Race is a social concept used historically to divide, control populations and reinforce social hierarchies. National Institutes of Health (NIH) promoted the use of racial and ethnic categories created by the US Census Bureau's Office of Management and Budget (OMB) in biomedical research contexts [1,2].They require the new approach toward precision medicine to look at the diversity further than just genomic biomarkers. Racial diversities may help facilitate a deeper understanding of clinical outcomes. This study aims to analyze the Electronic Medical Record (EMR) to find the possible effect of different races on the diagnosis and treatment of CVD.

2. Method

2.1. Study Cohort and Data Selection

The data is from de-identified EMR from 960,129 patients admitted to UCSF during 2011-2018. After authorization to access EMR data for research, the following cohort search criteria were developed for Coronary Artery Disease (CAD), commonly referred to as Ischemic Heart Disease (IHD), based on the ICD10 code (120-125). Patients with missing values specifically for the ICD10 code were excluded. Patients defined as unknown and unspecified were excluded. Patients who met the above criteria led to a cohort size of ~33,000 patients with CAD.

All data are extracted with MySQL queries from original datasets at UCSF medical center. Characteristics are shown in Table 1. Our data shows five different races in Table 2. The patients whose medical history does not include at least one element from the Current Procedural Terminology (CPT) codes that we defined as various cardiac procedures such as therapeutic and diagnostic catheters, various lab tests, and CABG were eliminated from the initial cohort of patients.

2.2. Data Analytics Platform and Setup.

We developed an algorithm that creates a sequence of events and finds the number of days between two events (e.g., the first electrocardiogram (EKG) test and Coronary artery bypass graft surgery (CABG)). Then, we iterated over the entire patient dataset and categorized these numbers of days by each patient's given race. After obtaining the average number of days between different events for each race (e.g., the average number of days between the first EKG and CABG for all races), we performed pairwise 2 sample T-tests corrected with Sidak correction. Lastly, the algorithm was able to take in a threshold value, and this threshold value helped us remove more possible outliers that may skew the results. This threshold represents the maximum number of days allowed in between events. Otherwise, we removed that patient. For example, if a patient has 500 days between their first EKG test and CABG and the threshold is set at 365, this patient is not considered when calculating the average amount of days. The next step was to search for evidence of race-based differences in different race groups. To validate this hypothesis, we determined the first point of an encounter with a physician when a patient was suspected of having CAD. We included

the treatment path with suspicion of potential CAD that combined both procedures and medications. Any medication that belongs to the classes of cardiovascular and cardiac drugs, antiplatelets, aspirin, beta-blockers, and statins is included as the starting point for medication. This medication is considered a “first event” for patients suspected of being at risk for CAD. Then, both the men and women datasets are merged on the same paths of the event for each race group defined by experts and ready for hypothesis validation based on p-value < 0.05 and p-value <0.00851 (after Bonferroni Correction) for significant differences. RStudio[2] and Python[3] in the Jupyter[4] notebook have been used for data characteristics and time-series sequence analysis for the path of events and finding possible differences in each race group.

3. Results

Our previous study found significant gender differences in the average waiting days from the first event to diagnostic catheterization [4]. We did not find significant gender-based differences between diagnostic catheterization and therapeutic cardiac catheterization (Percutaneous Coronary Intervention-PCI). We did not find any significant differences between diagnostic catheterization and CABG as well. Since this project requires a pairwise comparison with each race group, we implemented a Bonferroni correction[5] method with the original threshold at p=0.05. Therefore, the new threshold set to p =0.00851. We added a Bonferroni to reduce the probability of type I errors while we performed multiple sample tests to compare different races. We compared different combinations for more than 70 race group combinations (e.g., American Indian or Alaska Native versus Black or African American). Given this situation, the results are as following. There are no significant differences between different races from first event to diagnostic catheterization with both p-value < 0.5 (original threshold) and p-value < 0.00851(after Bonferroni correction). There are no significant differences from the first event to PCI between different races. There are differences between the first event to CABG, as shown in Table 3 between Asian versus American Indian or Alaska Native, American Indian or Alaska Native versus White or Caucasian. From the first event to CABG, Asians waited shorter than other races. White or Caucasians also waited shorter than American Indian or Alaska Natives. There are no significant differences in pairwise comparison for this category after applying correction with the 0.00851 thresholds instead of 0.05 for p-value. According to a study by Garcia et al., the black population in the United States has worse cardiovascular health and higher cardiovascular morbidity and mortality rates than other racial groups [3]. Our study shows that Native Americans or Black have waited longer versus Asian and white to get the invasive procedure from diagnostic catheterization. That could be one of the reasons for worse outcomes in this population. American Indian or Alaska Native also waited longer than any other race groups. Figure 1 shows the average time between each pair of events for different races.

[2] https://github.com/rstudio/
[3] https://www.python.org
[4] https://jupyter.org
[5] https://en.wikipedia.org/wiki/Bonferroni_correction

4. Discussion

This finding opens several hypotheses, such as considering Black and African Americans sooner to surgery if it is needed or convince them for surgery; look up the social, economic, environmental reason for longer waits. Also finding the comorbidities and other related illnesses for individual races with related social characteristics plus genetics is essential for prevention and a better treatment plan.

5. Limitation

The biggest limitation is considering just EMR for this study. There is an urgent need for considering other resources such as Social Media for racial discrepancies and the possible reasons.

6. Conclusion

Data analytics over UCSF patients who had an invasive procedure makes a few questions, hypotheses, and roadmaps for the future research. For example, what are the possible reasons for the worse outcome in Black or African American population? Maybe staying longer to go for PCI and/or CABG could be one of the reason. Or what is the reason for Asian population for shorter waiting time? This could reflect the effect of culture, environment and social economic in San Francisco BAY Area. More study is needed to confirm this new hypothesis. According to the American college of cardiology, solving racial health disparities is an essential and pressing priority for all in health care. There is an urgent need to address these, both locally and nationally [5]. We plan to extend our data set to UC System data, to study this hypothesis and discover more knowledge for the effect of races in CVD diagnosis and treatment and health outcome.

References

[1]. Callier SL. The Use of Racial Categories in Precision Medicine Research. Ethn Dis. 2019 Dec 12;29(Suppl 3):651–658. doi: 10.18865/ed.29.S3.651. [PubMed: 31889770]

[2]. Charkoudian N, Fadel PJ, Robinson AT, Zucker IH. Call for papers on racial differences in cardiovascular and cerebrovascular physiology. Am J Physiol Heart Circ Physiol. 2020 Aug 1;319(2):H249–H250. doi: 10.1152/ajpheart.00524.2020. Epub 2020 Jul 3. [PubMed: 32618515]

[3]. Garcia M, Almuwaqqat Z, Moazzami K, Young A, Lima BB, Sullivan S, Kaseer B, Lewis TT, Hammadah M, Levantsevych O, Elon L, Bremner JD, Raggi P, Shah AJ, Quyyumi AA, Vaccarino V. Racial Disparities in Adverse Cardiovascular Outcomes After a Myocardial Infarction in Young or Middle-Aged Patients. J Am Heart Assoc. 2021 Sep 7;10(17):e020828. doi: 10.1161/JAHA.121.020828. Epub 2021 Aug 25. [PubMed: 34431313]

[4]. Panahiazar M, Bishara AM, Perri J. A Trend of Eight-Years Big Data Analytics of Electronic Medical Records to Review and Study Diagnosis and Treatment of Coronary Artery Disease in Different Genders, Res. Sq (2021) 1–16. doi:10.21203/rs.3.rs-366930/v1.

[5]. Suhendra AD, Asworowati RD, Ismawati T. Racial Differences in Long-Term Cardiovascular Outcomes. JACCg. 2020;5:43–54.

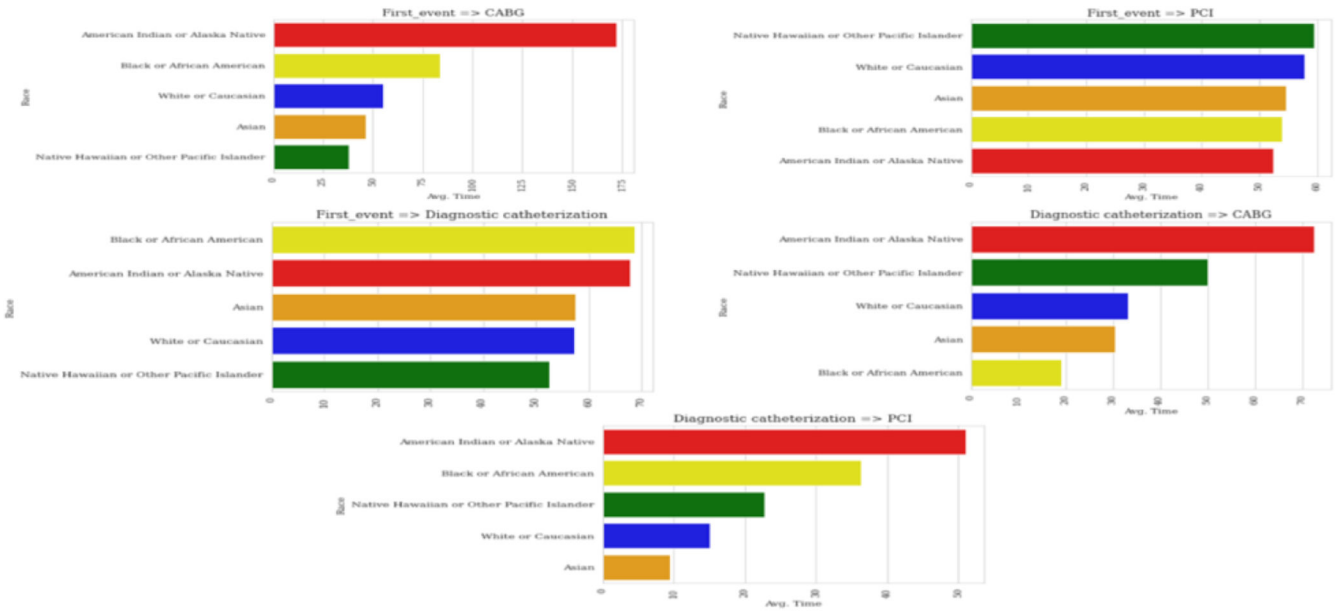

Figure 1.
Average waiting days (time) between each event for different races.

Table 1.

It shows the characteristics for a total of patients with CAD diseases (ICD10 code I20-I25).

Characteristics	Percentage	Characteristics	Average
Never Smoked	58.71	Age	69 (y)
Passive Smoke Exposure	0.58	BMI	31.74
Former Smoker	37.8	Blood Pressure (Diastolic)	69.95
Current Every day Smoker	5.25	(Systolic)	130.4
Ethnicity (Hispanic or Latino)	9.42	Cholesterol (Total)	168.65
Ethnicity (Not H or L)	82.71	Cholesterol (LDL)	91.90
Stroke	35.15	Cholesterol (HDL)	52.07
Status (alive)	88.79		
Status (deceased)	10.21		

Table 2.

It shows the racial distribution for women and men.

Race	Women	Men	Total
Asian	2126	3019	5145
American Indian or Alaska Native	73	110	183
Black or African American	1243	1403	2646
Native Hawaiian or Other Pacific Islander	269	377	646
White or Caucasian	6124	11649	17773

Table 3.

A comparison between different race groups shows with the p-value <0.05 with 95% confidence interval and p-value < 0.00851 after correction. We show the group of races that are significant with original or corrected threshold. diagnostic catheterization (d-Cath), Versus (v), nonsignificant(nS), Significant (S) Average Waiting Days(AWD), p-value(p)

Pair of event	Race	p, <0.00851, < 0.05	AWD
first-event => CABG	Asian v American Indian or Alaska Native	0.02015, nS, S	46 v 172
	Asian v Black or African America	0.04713, nS, S	46 v 83
	White or Caucasian v American Indian or Alaska Native	0.04044, nS, S	55 v 172
d-Cath=>.PCI	Asian v American Indian or Alaska Native	0.0081, S, S	9 v 51
	Asian v Black or African American	0.0060, S, S	9 v 36
	White or Caucasian v Black or African American	0.0195, nS, S	15 v 36
d-Cath => CABG and first-event =>PCI and first-event =>PCI and first-event => d-Cath	No Significant differences between any races		