Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

________________________________ _______________

Huige Jiang Date

Investigation of relationship between psychosocial profile and cardiovascular health among African Americans

By

Huige Jiang

Degree to be awarded: Master of Science in Public Health.

Department of Biostatistics and Bioinformatics

____

Yi-An Ko, PhD

Thesis Advisor

____

Mary Kelley, PhD

Reader

Investigation of relationship between psychosocial profile and cardiovascular health among African Americans

By

Huige Jiang

Bachelor of Science

Georgia Institute of Technology

2017

Thesis Committee Chair: Yi-An Ko, PhD

An abstract of

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Biostatistics

2020

Abstract

Investigation of relationship between psychosocial profile and cardiovascular health among African Americans

By Huige Jiang

Background: Neighborhood and psychosocial health have known association with cardiovascular health and African Americans have been identified as a population with high risk for cardiovascular disease. Previous studies have shown that cardiovascular health is affected by factors such as depression and neighborhood conditions among different populations. However, a thorough study that considers multiple dimensions of neighborhood perceptions and psychosocial health in the African American population has not been conducted.

Methods: We examined a sample of 502 African Americans in the metropolitan area from the Morehouse Emory Center for Health Equity (MECA) Study. We first identified clusters of samples based on their neighborhood perception and psychosocial profile using four unsupervised clustering methods. We then compared differences in arterial stiffness (measured by pulse wave velocity and augmentation index) among clusters to investigate underlying relationships between psychosocial well-being and their vascular function, as a measure of subclinical cardiovascular disease. We also used four supervised machine learning methods to identify the most significant factors affecting arterial stiffness.

Results: Clustering analysis results show that subjects who are psychosocially healthier have better cardiovascular health, indicated by lower augmentation index. However, pulse wave velocity is not associated with psychosocial profile. Additionally, different arterial stiffness measures are mainly associated with slightly different psychosocial factors.

Conclusion: Neighborhood perceptions and psychosocial profile are associated with arterial stiffness, measured by augmentation index among African Americans. Factors such as optimism and environmental mastery play a major role in affecting subclinical cardiovascular health.

Investigation of relationship between psychosocial profile and cardiovascular health among African Americans

By

Huige Jiang

Bachelor of Science

Georgia Institute of Technology

2017

Thesis Committee Chair: Yi-An Ko, PhD

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Biostatistics

2020

Introduction

Cardiovascular disease (CVD) is defined as a group of disorders of heart and blood vessels.[1] Cardiovascular disease includes coronary heart disease (CHD), stroke, peripheral vascular disease, congenital heart disease, endocarditis, and many other conditions.[2] Common CVDs include heart attacks and strokes, which occur when oxygenated blood flow to the heart or the brain is blocked. [3] CVDs are the top leading cause of death globally and an estimated 17.9 million people died from CVDs in 2016, representing 31% of all global deaths. Studies have shown that over three quarters of CVD deaths take place in low- and middle-income countries[4] and most CVDs can be prevented by addressing behavioral risk factors such as tobacco use, unhealthy diet and harmful use of alcohol.[1]

Cecelja et al. (2012) found that arterial stiffness is an independent predictor of cardiovascular morbidity and mortality and it is typically measured using pulse wave velocity (PWV) and augmentation index (AIX).[5] PWV is the measure of the speed of arterial pressure waves traveling along the aorta and large arteries, whereas AIX is derived from the ascending aortic pressure waveform, and both measurements are inversely related to cardiovascular health. [6]

While numerous studies have identified the significance of psychosocial factors for the development of CVD, few studies have focused on significant psychosocial factors among the African American population, which has higher CVD mortality rates than other ethnic groups in the United States.[7] Furthermore, in most previous studies, a specific category of the psychosocial factors was examined, such as depression, marital stress and job stress. Such approaches could be potentially limited in that some significant factors are not considered. For example, Horsten et al. (2000) discovered that the presence of two or more depressive symptoms and lack of social integration independently predicted recurrent cardiac events in women with coronary heart disease;[8] Wulsin et al. (2005) conducted an analysis based on the Framingham Heart Study, which included US adults (mostly Caucasians) aged 18–77 and their findings underscored the importance of further research into the pathogenesis and prevention of excess mortality experienced with depressive symptoms;[9] Lee et al. (2003) examined the Nurses' Health Study among 54,412 women aged 46 to 71 years who were registered nurses and arrived at the conclusion that High levels of care provision to grandchildren (and possibly children) may increase the risk of CHD among women[10]. Felix et al. (2019) reported that stressful life events are related to CVD whereas resilience is not among older African American women.[11]As can be seen, few of these studies were based on an African American cohort and most of them examined one single specific aspect of psychosocial factors. In our study, we consider an extensive list of relevant psychosocial factors in multiple categories simultaneously, including depression, neighborhood wellness, social support, etc. Therefore, we offer some more up-to-date machine learning strategies to analyze and discover relationships between psychosocial factors and CVD risk in African Americans.

In this study, we use a dataset of 502 African Americans with their psychosocial and demographic information. The analysis will begin with a cluster analysis to identify clusters of subjects based on personal wellness and neighborhood wellness factors; the purpose of this step is to examine if there is significant difference in arterial stiffness between these clusters, which could be an indication of a crucial relationship between psychosocial factors and vascular health. Consequently, several machine learning methods will be applied to identify the most significant psychosocial predictors of vascular health using the two measures of arterial stiffness as outcome variables, respectively.

Methods

Study Participants

The data for our study was obtained from the Morehouse-Emory Cardiovascular Center for Health Equity Study (MECA). The goal of the MECA study was to identify neighborhoods where cardiovascular disease (CVD) rates among African Americans are higher and lower than average as well as to investigate characteristics associated with resilient and at-risk areas.[12] For our study, we obtained individual-level data on personal wellness and neighborhood wellness of 502 African Americans residing in Atlanta metropolitan area from the MECA study. Nearly all 502 subjects were included in the unsupervised cluster analysis. However, for any subsequent analysis associated with the outcome variables (PWV and AIX), we only used a subset of the 502 subjects. This is due to some unavoidable loss of samples throughout the analysis process, which is depicted in Figure 1. As an example, for k-means only 427 subjects were included for subsequent outcome-related analysis, and the remaining 75 subjects were excluded. The demographic information, education level, and health-related characteristic information (e.g., smoking habits) are summarized and compared in Table 1 between the subjects included in the analysis and those who were excluded. We used the same 427 subjects for supervised machine learning analysis.

Psychosocial Factors: Personal Wellness and Neighborhood Wellness

Domains of personal wellness and neighborhood wellness were assessed via telephone survey in 55 subjects between August 2016 and October 2016 and in-person visit in 477 subjects between August 2017 and July 2019.[14] The variables selected included factors that previous studies have shown to be associated with cardiovascular health. More specifically, discrimination[15], emotional abuse, environmental mastery, traumatic experiences, physical punishment, optimism[16,17], life purpose[18,19], resilience towards difficulties[20], forced sexual experiences, depression level[21], social support and religious practices were selected as measures of personal wellness. Social support was evaluated based on responses from survey questionnaire using an eight-item modified Medical Outcomes Study Social Support Survey (mMOS-SS) method proposed by Moser et.al (2012).[23] Activity, aesthetic quality, cohesion, healthy food access, safety, violence, walking environment were selected to assess neighborhood wellness.[22] These information are summarized in Table 2.

Vascular Function Variables

The outcome variables in our study are carotid-femoral artery pulse wave velocity (PWV), which was determined using transcutaneous Doppler flow velocity recordings simultaneously over the common carotid artery and the femoral artery, and augmentation index (AIX), a composite measure of the magnitude of arterial wave reflections and systemic arterial stiffness.[13] A total of 463 out of the 502 subjects had complete outcome measures recorded. PWV is found to be affected by systolic blood pressure, age and gender whereas AIX is affected by systolic blood pressure, age, gender and height. Therefore, PWV and AIX were adjusted for these factors, respectively using linear regression. The residuals of linear regression were extracted for the subsequent analysis.

Statistical analysis

Unsupervised cluster analysis

We first conducted cluster analysis to identify groups of subjects with similar psychosocial conditions. Four clustering techniques are considered- k-means, partition around medoids (PAM), divisive hierarchical clustering and agglomerative hierarchical clustering

1) K-means

K-means clustering algorithm is based on iterative refinement. To begin with, k samples are randomly selected as cluster centers (centroids). Next, all other samples are assigned to the nearest centroid, where nearest is defined based on Euclidean distance. After centroid assignment, each cluster is "updated" to be its mean and every sample is re-examined to be assigned to the newly closest centroid. This process is repeated until cluster assignment for all samples remain constant. [24]

2) PAM

PAM is a more robust alternative for k-means clustering. The algorithm starts by selecting k samples to be medoids and assigning all other samples to the nearest medoid. Next, each selected medoid and non-selected sample is swapped if the sum of dissimilarities could be decreased. This process is iterated until the sum of dissimilarities cannot be reduced. [25] One advantage of PAM is that it handles ordinal variables easily. Due to the way our data was collected, a number of variables in our study could be reasonably treated as ordinal and therefore PAM is a better choice in this regard.

3) Divisive hierarchical clustering

Divisive hierarchical clustering works from top to bottom. It begins with one single cluster with all samples, and at each iteration step, the most heterogeneous cluster is divided into two. The iteration continues until each sample is in its own cluster.[26]

4) Agglomerative hierarchical clustering

In contrast, agglomerative hierarchical clustering works from bottom to top. Initially, each sample is regarded as a single cluster. Then at each step of the iteration, two of the clusters that are most similar are combined to become a new big cluster. The procedure continues until all samples are included in one single cluster. [26]

Selecting optimal number of clusters

To aid the clustering analysis process, the optimal number of clusters is first determined. We consider three methods here: elbow method, Silhouette method and gap statistic method. Each method will have its own optimal number of clusters, and we select the number which the majority of methods decide on. The elbow algorithm plots the total within-cluster sum of square over a range of possible numbers of clusters, and the location of a bend (elbow) in the plot is considered as an indicator of the appropriate number of clusters; the Silhouette method utilizes the concept of average silhouette score, which is an indicator of how well each subject lies within a cluster, and picks the optimal number of clusters that maximizes average silhouette score; the gap statistic method picks the optimal number of clusters that makes the clustering structure most far away from uniform distribution of points generated using Monte Carlo simulations of the sampling process.[24]

Once the clusters are determined using each method, we examined the differences of each personal wellness and neighborhood wellness variable; we processed variables to make all variables to be "in the same direction" so that higher values indicate better psychosocial condition. Additionally, we compared adjusted PWV and adjusted AIX between clusters by two-sample t-test.

Remarks

Among the four clustering methods considered, k-means and PAM clustering require the user to pre-specify the number of clusters while hierarchical clustering does not. Since k-means and hierarchical clustering considered handle only continuous variables, frequency of religious practices was excluded during the clustering analysis process and all continuous variables were standardized to have a mean of 0 and standard deviation of 1. Unlike the other measures, PAM does not require all continuous measures but rather allows for various types of categorical variables, including ordinal as well as symmetric and asymmetric nominal variables. Therefore all psychosocial variables were included in the PAM analysis. This is important for the current data, as the psychosocial measures are predominantly ordinal in nature and therefore have restricted variance. While clusters are identified in a parallel fashion in k-means and PAM, hierarchical clustering builds clusters incrementally, resulting in a tree like structure or a parent child relationship among clusters. Consequently, hierarchical clustering is most applicable when the underlying data has a hierarchical structure and the recovery of the hierarchy is desired. On the contrary, k-means has the advantage of dealing with datasets with a large number of variables and is less sensitive to outliers.[27] Since k-means and PAM start with random selection of samples as centroids, results might not be consistent for multiple runs unless a seed is predetermined; on the contrary, results from hierarchical clustering are reproducible.

Supervised machine learning analysis

In order to identify some of the most crucial personal wellness and neighborhood wellness features that have the most significant impact on arterial stiffness, we also applied supervised machine learning algorithms to explore the relationships. Here we considered random forest, support vector machine (SVM), and relevance vector machine (RVM).

1) Random forest

Random forest is an ensemble of decision trees: each decision tree produces its own prediction and the class with the most votes serve as the final prediction of random forest; it utilizes bagging and feature randomness when building each single tree to create an uncorrelated forest of trees whose prediction more accurate than that of any individual tree. [28]

2) SVM

The objective of SVM is to find a hyperplane in an N-dimensional space (N is the number of features) that separates the data points as much as possible. In the SVM algorithm, we aim to maximize the margin between the data points and the hyperplane. The loss function helps maximize the margin and a regularization parameter is added to balance the margin maximization and loss.[29]

3) RVM

RVM is a Bayesian treatment of a generalized linear model of identical functional form to the SVM and provides probabilistic classification.[30]

Measures of importance

For random forest, we considered three different measures of importance, mean decrease in accuracy, mean decrease in MSE and average absolute deviation from the median (AAD) produced by 1 dimensional sensitivity analysis (1D-SA). Mean decrease in accuracy is measured as a result of variable being permuted; mean decrease in MSE relates to the loss function; 1D-SA examines the impact of each feature by observing the change in the model's outcome as we change the feature value. For SVM and RVM, we only consider one type of importance measure: AAD produced by 1D-SA.

Remarks

Different from most of the clustering analysis methods, all three supervised machine learning methods are able to handle both continuous and categorical variables simultaneously, therefore all psychosocial factors are included. Furthermore, variables are allowed to be on various scales and do not need to be standardized. Machine learning methods are usually used to make predictions in either classification or regression, but here we only use them to rank feature importance. Random forest performs better in situations with large data sets but results might not be consistent for multiple runs with different seeds. On the contrary, SVM performs better in cases where number of dimensions is greater than the sample size.

Results

The study cohort consisted of 502 subjects with the majority of psychosocial data. As described in Methods, only 427 subjects were assigned a cluster and had complete outcome data and therefore were included for subsequent outcome-related analysis, and the remaining 75 subjects were excluded. Table 1 summarizes these characteristics for each group and compares the differences between the two groups. Compositions of these two groups are similar for alcohol consumption in past 30 days, smoking status, annual household income, average PWV, age and household size; however, there are significant differences in gender, education level, diabetes status, hypertension status, high cholesterol status and AIX. For example, males comprised 40.5% of the samples in the larger group (N=427) whereas only 21.3% in the smaller group (N=75).

Unsupervised machine learning analysis

Since preliminary analysis indicated that the optimal number of clusters for all 4 clustering methods is two, we grouped all subjects into 2 clusters using each method. Cluster assignment of subjects are slightly different based on the 4 clustering methods. Cluster-specific means for all the 18 psychosocial variables were used to describe the characteristics for each cluster profile (Table 3). Although the 4 methods did not cluster subjects exactly same, cluster 1 in general has better psychosocial health indicators, while cluster 2 has lower psychosocial health indicators. Additionally, we also compared the average PWV (adjusted for systolic blood pressure, age and gender) and average AIX (adjusted for systolic blood pressure, age, gender and height) between the 2 clusters. Relevant results are summarized in Figure 2. Shown in Figure A1-A4, the adjusted average PWV of the 2 clusters are similar. Upon conducting a student t-test, the p-values for difference in adjusted PWV between the 2 clusters based on k-means, PAM, agglomerative hierarchical clustering and divisive hierarchical clustering are 0.77, 0.74, 0.89 and 0.96, respectively, showing no significant difference. In contrast, adjusted AIX appears to be different between the 2 clusters. Shown in Figure B1-B4, the adjusted AIX of cluster 1 is generally lower than cluster 2, indicating better vascular function. Upon conducting a student t-test, the p-values for difference in adjusted AIX between the 2 clusters based on k-means, PAM, agglomerative hierarchical clustering and divisive hierarchical clustering are 0.034, 0.032, 0.022 and 0.076, respectively, showing mildly to strongly significant difference. This implies that AIX is associated with psychosocial status whereas PWV is not, after adjusting for covariates. Specifically, better psychosocial well-being is associated with better vascular function as seen in the two clusters.

Supervised machine learning analysis

Supervised machine learning analysis provides further investigation towards variable importance. Shown in Figure 3A-5A, based on adjusted average PWV, the top 4 significant factors are environmental mastery, optimism, depression and neighborhood walking environment were ranked highest by random forest; optimism, forced sexual event, emotional abuse and neighborhood activity by SVM; social support, depression, neighborhood cohesion and neighborhood safety by RVM. Shown in Figure 3B-5B, based on adjusted AIX, the top 4 significant factors are neighborhood safety, optimism, environmental mastery and purpose in life are ranked higher by random forest; purpose in life, environmental mastery, optimism and religious practice by SVM; neighborhood cohesion, depression, neighborhood safety and optimism by RVM. As can be seen, variable importance ranking is outcome driven and also depends on selection of methods. Nevertheless, optimism, environmental mastery, neighborhood safety seemed to be the most commonly top ranked factors related to arterial stiffness.

Discussion

Our results have shown that between the two arterial stiffness measures, PWV and AIX, AIX seems to be closely associated with psychosocial well-being. Each arterial stiffness measure is mainly associated with a particular set of psychosocial factors. More specifically, based on the supervised learning results, adjusted PWV is most associated with optimism and depression; adjusted AIX is most related to optimism, environmental mastery and purpose in life.

The results from our study are quite consistent with the findings from previous studies. Optimism has been reported to alleviate the risk of CVD[31]; Depression has long been shown to largely affect cardiovascular health[8]. In addition, our results offer some novel perspectives on the relationship between particularly arterial stiffness and psychosocial well-being.

Our study has a number of limitations. First, there were differences in sex, education level, diabetes status, hypertension status, high cholesterol status and AIX between the subjects included in the analysis and those who were excluded. This gap could potentially cause bias in our results, therefore statistical correction in the future is much needed. Possible approaches include stratification and imposing weights to various samples. Second, all eligible variables were considered in the unsupervised cluster analysis and no preliminary feature selection step was involved. This might bring noise to the cluster assignment process. Currently, there is no standardized feature selection methods for the four methods in our study. Future work includes the development of effective feature selection procedures. Third, small sample size and lack of validation of the results from cluster analysis imposes uncertainty to our conclusions. Fourth, this was a cross-sectional observational study. The observed association may not indicate a causal relationship. Lastly, we treated some psychosocial variables as continuous in our study for k-means and hierarchical clustering, causing a larger separation of clusters that might not reflect the truth, since processing variables as continuous deflates the variance when there are many ties, which could be a false premise. In comparison, when the same set of psychosocial variables are treated as ordinal for PAM, separation is smaller.

In conclusion, the results from our study shed light on the relationship between cardiovascular health and psychosocial well-being among African Americans on various levels. First, multiple clustering analysis strategies confirmed that clusters based on psychosocial well-being have notable distinctions in arterial stiffness. Second, compared to PWV, AIX seems to be much more closely connected to psychosocial well-being. Third, our results are in agreement with previous findings in that cardiovascular health is associated with various psychosocial factors. By applying multiple supervised machine learning analysis, we identified that PWV is mainly related to optimism and depression; AIX is mostly related to optimism, environmental mastery and purpose in life. These findings could potentially guide the intervention and prevention of CVD. Based off what we found so far, future research could be conducted to explore if varying significant psychosocial factors within patients could improve their cardiovascular health, as well as the underlying interactions among these factors, which would serve as a further step into CVD intervention.

Tables and Figures

Table 1. Distribution of the Baseline Characteristics (Discrete variables are described with N and %; normally distributed continuous variables are described with mean and standard deviation; non-normal continuous variables are described as median, 25% and 75% quartiles) for the subjects included in the analysis (N=427) and those excluded from the analysis (N=75).

	Subjects included in the analysis (N=427)	Subjects excluded from the analysis (N=75)	P-value*
Gender			
Male	173 (40.5)	16 (21.3)	0.0016
Female	254 (59.5)	59 (78.7)	
Education attainment			
Elementary school	4 (0.9)	3 (4.2)	0.018
Some high school	27 (6.3)	6 (8.3)	
High school graduate	97 (22.7)	13 (18.1)	
Some college or technical school	155 (36.3)	29 (40.3)	
College graduate	144 (33.7)	21 (29.2)	
Alcohol consumption in past 30 days			
Never	217 (51.5)	39 (52)	0.99
1-2 times/week	47 (11.1)	9 (12)	
3-4 times/week	14 (3.3)	2 (2.7)	
5-6 times/week	4 (0.94)	0 (0)	
Daily	9 (2.1)	1 (1.3)	
Once/month	74 (17.4)	14 (18.7)	
2-3 times/month	60 (14.1)	10 (13.3)	
Smoking status			
Current smoker	98 (23)	18 (24)	0.8303
Quit smoking < = 12 months ago	18 (4.2)	4 (5.3)	
Never smoked/quit > 12 months ago	311 (72.8)	53 (70.6)	
Annual household income			
Less than $10,000	92 (22.7)	13 (19.4)	0.5769
$10,000 to less than $15,000	40 (9.9)	10 (14.9)	
$15,000 to less than $20,000	34 (8.4)	7 (10.4)	
$20,000 to less than $25,000	35 (8.6)	5 (7.5)	
$25,000 to less than $35,000	63 (15.5)	6 (9.0)	
$35,000 to less than $50,000	53 (13.1)	9 (13.4)	
$50,000 to less than $75,000	53 (13.1)	10 (14.9)	
$75,000 or more	36 (8.9)	7 (10.4)	
Diabetes status			
Yes	105 (25.2)	28 (37.8)	0.0051
No	314 (75.3)	46 (62.2)	
Hypertension status			
Yes	225 (52.9)	50 (69.4)	0.0021
No	200 (47.1)	22 (30.6)	
High cholesterol status			
Yes	155 (36.6)	40 (9.5)	0.00040
No	268 (63.4)	33 (7.8)	
	Mean (SD)	Mean (SD)	
Average pulse wave velocity	7.60 (2.09)	7.54 (1.50)	0.91
Augmentation index adjusted for heart rate	19.96 (12.98)	23.95 (9.84)	0.035
	Median (Q1, Q3)	Median (Q1, Q3)	

Age	56 (47,61)	56 (49,61)	0.90
Household size	2 (1,3)	2 (1,4)	0.17

*P-values for comparing the two groups were obtained using two-proportions z-test for categorical variables and 2-sample t-test for continuous variables

Table 2. Summary of the Baseline Psychosocial Information (discrete variables are described with N and %; continuous variables are described as median, 25% and 75% quartiles) in the study subjects (N=502)

Psychosocial factor (possible range)	Median	(Q1, Q3)
Discrimination score (1-4)	1.5	(1.1, 2.0)
Emotional abuse score (0-1)	0.2	(0.0, 0.6)
Total environmental mastery score (1-6)	4.8	(4.0, 5.4)
General traumas score (0-1)	0.4	(0.3, 0.5)
Purpose in life score (1-6)	5.1	(4.4,5.6)
Neighborhood activity score (1-4)	2.6	(2.0, 3.2)
Neighborhood aesthetic quality score (1-5)	3.8	(3.2, 4.4)
Neighborhood cohesion score (1-5)	3.5	(3.0, 4.4)
Neighborhood healthy food access score (1-5)	3.3	(2.6, 4.0)
Neighborhood safety score (1-5)	3.3	(2.7, 4.0)
Neighborhood violence score (1-4)	1.3	(1.0, 1.8)
Neighborhood walking environment score (1-5)	3.8	(3.2, 4.3)
Optimism score (1-5)	4.2	(3.5, 4.7)
Physical punishment score (0-1)	0.4	(0.0, 0.6)
Resilience score (0-4)	3.2	(2.7, 3.7)
Forced sexual event score (0-1)	0.0	(0.0, 0.3)
Depression score (0-63)	5.0	(2.0, 11.0)
Social support score (0-100)	12.5	(0.0, 34.4)
	N	%
Religious practices frequency		
At least weekly	184	37.0
At least monthly	169	34.0
At least yearly	110	22.1
Never	35	7.0

Table3. Summary of psycho-social and neighborhood perception variables of 2 cluster by 4 different clustering analysis methods in the form of mean (standard deviation).

Methods / Variables	K-means		PAM		Agglomerative hierarchical clustering		Divisive hierarchical clustering	
	Cluster 1	Cluster 2	Cluster 1	Cluster 2	Cluster 1	Cluster 2	Cluster 1	Cluster 2
Discrimination score (1-4) **1=more discrimination** **4=less discrimination**	3.28(0.5)	3.09(0.6)	3.31(0.48)	3.01(0.62)	3.42(0.39)	3.03(0.6)	3.3(0.49)	2.79(0.61)
Emotional abuse score (0-1) **0=more abuse** **1=less abuse**	0.66(0.37)	0.64(0.38)	0.72(0.36)	0.6(0.38)	0.9(0.21)	0.53(0.38)	0.75(0.34)	0.41(0.36)
Total environmental mastery score (1-6) **1=less mastery** **6=more mastery**	4.68(0.88)	4.61(0.84)	4.83(0.89)	4.39(0.74)	5.19(0.59)	4.31(0.82)	4.9(0.75)	3.8(0.61)
General traumas score (0-1) **0=more trauma** **1=less trauma**	0.62(0.21)	0.6(0.23)	0.62(0.21)	0.59(0.23)	0.66(0.21)	0.57(0.22)	0.64(0.21)	0.51(0.22)
Purpose in life score (1-6) **1=less purpose** **6=more purpose**	4.92(0.86)	4.86(0.85)	5.05(0.8)	4.66(0.87)	5.42(0.46)	4.56(0.87)	5.12(0.73)	4.11(0.8)
Neighborhood activity score (1-4) **1=less activity** **4=more activity**	2.51(0.82)	2.46(0.81)	2.63(0.8)	2.35(0.8)	2.73(0.79)	2.38(0.8)	2.66(0.78)	2.03(0.74)
Neighborhood aesthetic quality score (1-5) **1=less aesthetic** **5=more aesthetic**	3.92(0.74)	3.66(0.79)	3.95(0.73)	3.55(0.78)	4.16(0.61)	3.56(0.78)	3.98(0.67)	3.12(0.76)
Neighborhood cohesion score (1-5) **1=less cohesion** **5=more cohesion**	3.55(0.71)	3.4(0.85)	3.63(0.76)	3.25(0.78)	3.81(0.7)	3.26(0.76)	3.68(0.67)	2.77(0.74)
Neighborhood healthy food access score (1-5) **1=less access** **5=more access**	3.38(1.08)	3.17(1.13)	3.45(1.06)	3.04(1.13)	3.71(0.99)	3.02(1.09)	3.51(1.01)	2.52(1.08)
Neighborhood safety score (1-5) **1=less safety** **5=more safety**	3.43(0.96)	3.3(0.93)	3.56(0.89)	3.08(0.95)	3.7(0.87)	3.14(0.93)	3.59(0.84)	2.57(0.86)
Neighborhood violence score (1-4) **1=more violence** **4=less violence**	3.68(0.5)	3.4(0.73)	3.65(0.54)	3.38(0.72)	3.71(0.51)	3.43(0.68)	3.69(0.48)	3.02(0.8)
Neighborhood walking environment score (1-5) **1=less environment** **5=more environment**	3.7(0.81)	3.66(0.75)	3.83(0.76)	3.47(0.76)	3.94(0.66)	3.51(0.8)	3.86(0.69)	3.04(0.73)
Optimism score (1-5) **1=less optimism** **5=more optimism**	4.05(0.75)	4.04(0.8)	4.14(0.78)	3.93(0.75)	4.51(0.48)	3.77(0.79)	4.19(0.7)	3.59(0.83)
Physical punishment score (0-1)	0.62(0.33)	0.57(0.33)	0.61(0.33)	0.57(0.33)	0.74(0.29)	0.51(0.33)	0.63(0.33)	0.51(0.34)

0=more punishment **1=less punishment**								
Resilience score (0-4) **0=less resilience** **4=more resilience**	3.15(0.66)	3.11(0.77)	3.2(0.74)	3.03(0.66)	3.45(0.44)	2.93(0.77)	3.28(0.63)	2.62(0.74)
Forced sexual event score (0-1) **0=more events** **1=less events**	0.8(0.29)	0.75(0.27)	0.86(0.26)	0.77(0.3)	0.93(0.14)	0.76(0.32)	0.87(0.23)	0.69(0.36)
Depression score (-6-63) **6=more depression** **63=less depression**	32.23(7.52)	32.12(8.18)	33.86(7.04)	30.21(8.35)	36.4(3.37)	29.81(8.69)	34.86(5.04)	23.81(9.43)
Social support score (0-100) **0=less support** **100=more support**	82.34(20.9)	77.05(24.31)	84.88(26.34)	72.14(12.5)	91.55(12.53)	72.6(24.4)	85.93(17.92)	59.38(24.88)

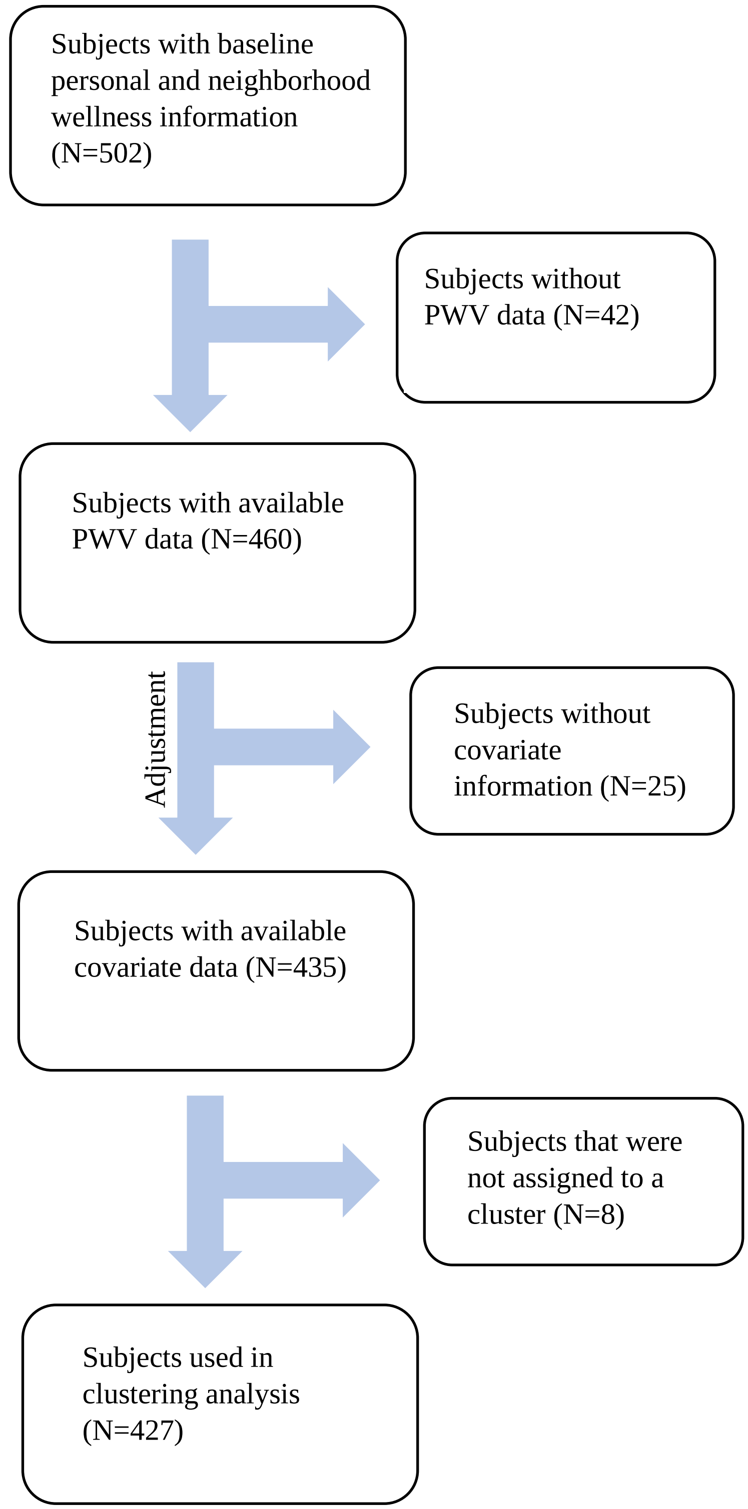


Figure1. Sample sizes in the study

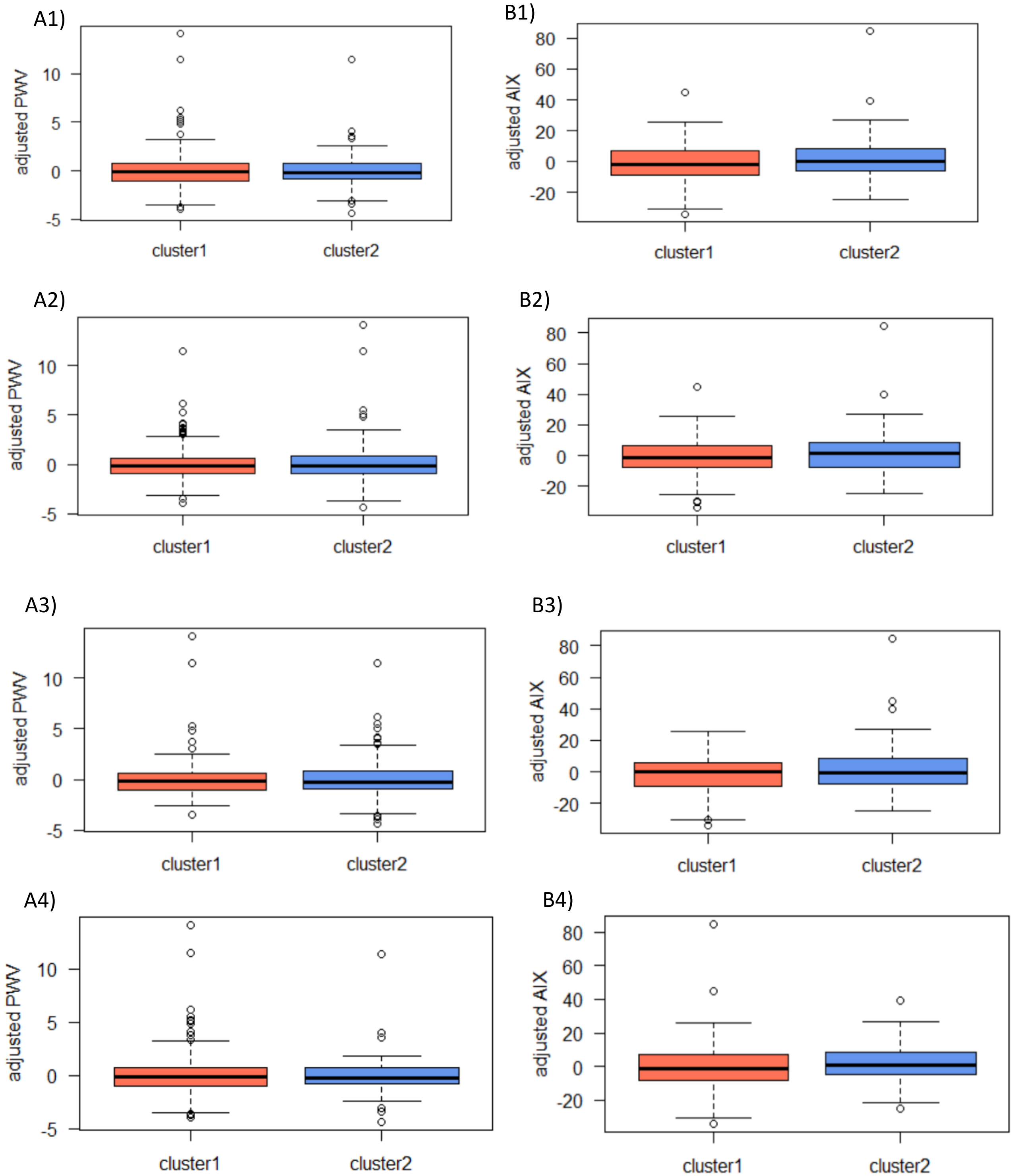
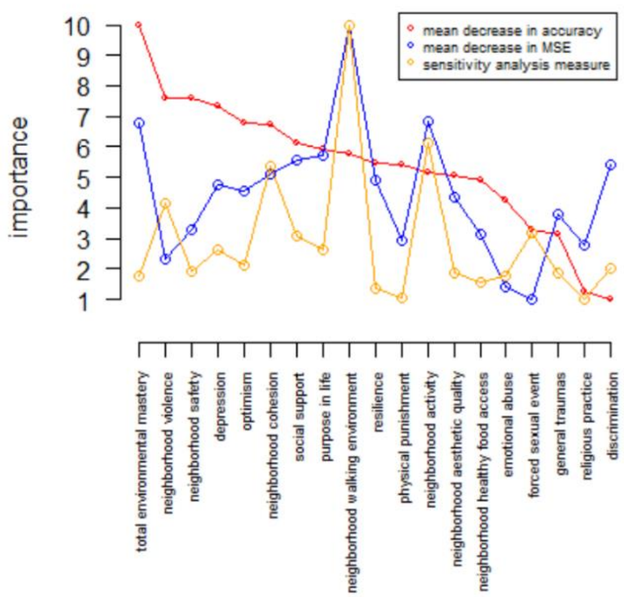
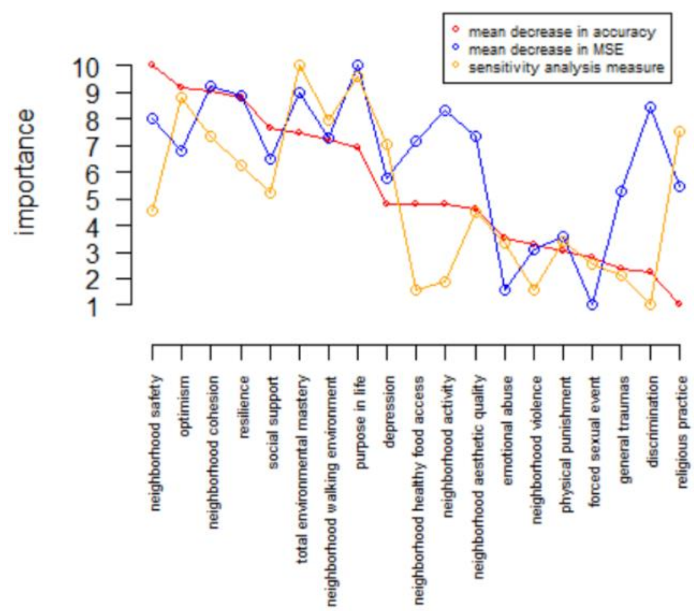

A1)
adjusted PWV
10
5
0
-5
cluster1
cluster2
B1)
adjusted AIX
80
60
40
20
0
-20
cluster1
cluster2
A2)
adjusted PWV
10
5
0
-5
cluster1
cluster2
B2)
adjusted AIX
80
60
40
20
0
-20
cluster1
cluster2
A3)
adjusted PWV
10
5
0
-5
cluster1
cluster2
B3)
adjusted AIX
80
60
40
20
0
-20
cluster1
cluster2
A4)
adjusted PWV
10
5
0
-5
cluster1
cluster2
B4)
adjusted AIX
80
60
40
20
0
-20
cluster1
cluster2

Figure 2. (A) Comparison of PWV adjusted for systolic blood pressure, age and gender; (B) Comparison of AIX adjusted for systolic blood pressure, age, gender and height of 2 clusters by 1) k-means 2) PAM clustering 3) agglomerative hierarchical clustering 4) divisive hierarchical clustering



(A)



(B)

Figure 3. Psychosocial profile importance rankings by random forest based on (A) average PWV adjusted for systolic blood pressure, age and gender (B) AIX adjusted for systolic blood pressure, age, gender and height

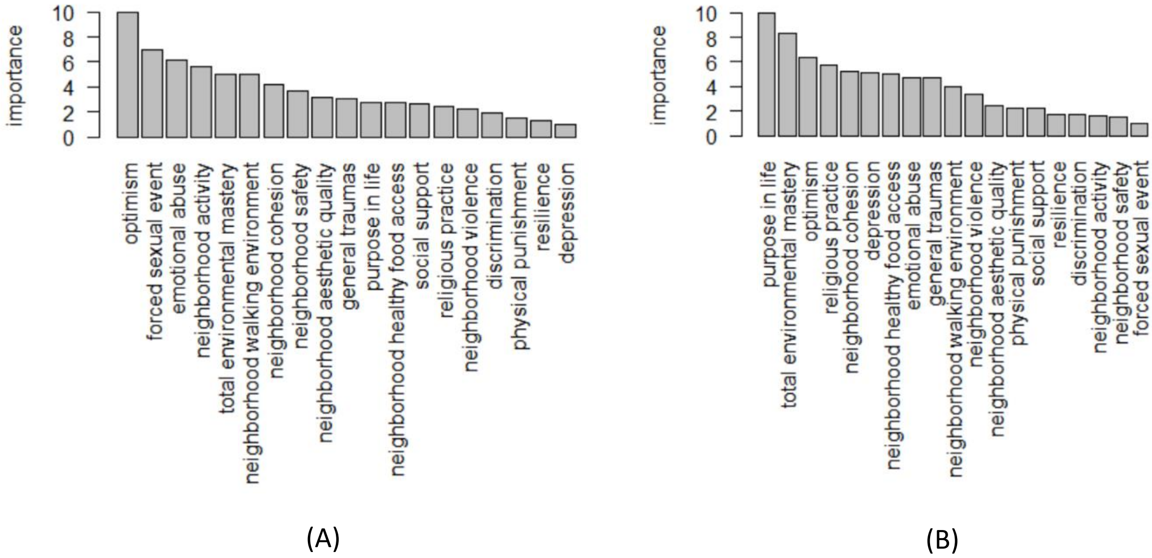
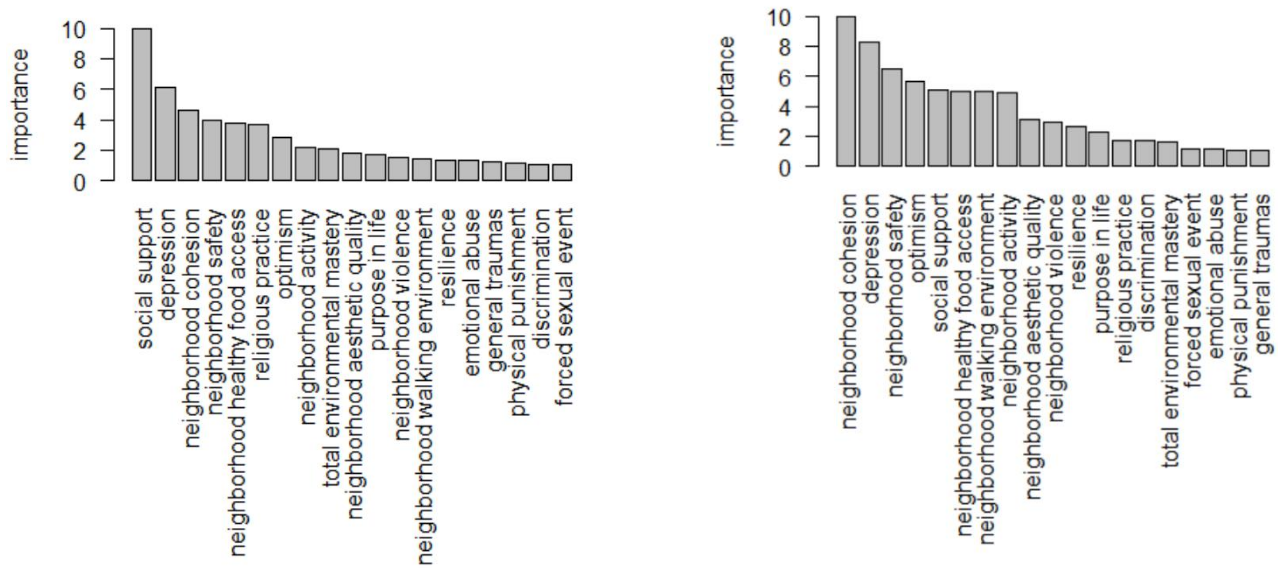


Figure 4. Psychosocial profile importance rankings by SVM based on (A) average PWV adjusted for systolic blood pressure, age and gender (B) AIX adjusted for systolic blood pressure, age, gender and height



(A) (B)

Figure 5. Psychosocial profile importance rankings by RVM based on (A) average PWV adjusted for systolic blood pressure, age and gender (B) AIX adjusted for systolic blood pressure, age, gender and height

References

1. Cardiovascular diseases (CVDs). (2020). Retrieved 3 February 2020, from https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)
2. Introduction Cardiovascular Disease, Introduction - Publication Bureau for Public Health, West Virginia. (1997). Retrieved 3 February 2020, from http://www.wvdhhr.org/bph/cvd/intro.htm
3. What is Cardiovascular Disease? (2017) Retrieved 3 February 2020, from https://www.heart.org/en/health-topics/consumer-healthcare/what-is-cardiovascular-disease
4. Gaziano TA, Bitton A, Anand S, Abrahams-Gessel S, Murphy A. Growing epidemic of coronary heart disease in low- and middle-income countries. Current problems in cardiology 2010;35(2):72-115.
5. Cecelja M, Chowienczyk P. Role of arterial stiffness in cardiovascular disease. JRSM cardiovascular disease 2012;1(4):cvd.2012.012016.
6. García, F. M., Gracia, A. D., Calaforra, O., Solaz, E., Vicente, A., Pichler, G., & Redon, J. (2019). Prognostic Value Of Changes In Arterial Stiffness In Essential Hypertension. *Journal of Hypertension*, *37*. doi: 10.1097/01.hjh.0000571904.50711.3f
7. Low CA, Thurston RC, Matthews KA. Psychosocial factors in the development of heart disease in women: current research and future directions. Psychosomatic medicine 2010;72(9):842-54.
8. Horsten M, Mittleman MA, Wamala SP, Schenck-Gustafsson K, Orth-Gomér K. Depressive symptoms and lack of social integration in relation to prognosis of CHD in middle-aged women. The Stockholm Female Coronary Risk Study. European heart journal 2000;21(13):1072-80

9. Wulsin LR, Evans JC, Vasan RS, Murabito JM, Kelly-Hayes M, Benjamin EJ. Depressive symptoms, coronary heart disease, and overall mortality in the Framingham Heart Study. Psychosomatic medicine 2005;67(5):697-702.

10. Lee S, Colditz G, Berkman L, Kawachi I. Caregiving to children and grandchildren and risk of coronary heart disease in women. American journal of public health 2003;93(11):1939-44.

11. Felix, A. S., Lehman, A., Nolan, T. S., Sealy-Jefferson, S., Breathett, K., Hood, D. B., … Williams, K. P. (2019). Stress, Resilience, and Cardiovascular Disease Risk Among Black Women. *Circulation: Cardiovascular Quality and Outcomes*, *12*(4). doi: 10.1161/circoutcomes.118.005284

12. Kim, J. H., Lewis, T. T., Topel, M. L., Mubasher, M., Li, C., Vaccarino, V., … Baltrus, P. T. (2019). Identification of Resilient and At-Risk Neighborhoods for Cardiovascular Disease Among Black Residents: the Morehouse-Emory Cardiovascular (MECA) Center for Health Equity Study. *Preventing Chronic Disease*, *16*. doi: 10.5888/pcd16.180505

13. The Morehouse-Emory Cardiovascular Center for Health Equity Study: Clinical Intervention Project - Full Text View. (n.d.). Retrieved from https://clinicaltrials.gov/ct2/show/NCT0330881

14. Topel, M. L., Kim, J. H., Mujahid, M. S., Ko, Y. A., Vaccarino, V., Mubasher, M., … Lewis, T. T. (2019). Individual Characteristics of Resilience are Associated With Lower-Than-Expected Neighborhood Rates of Cardiovascular Disease in Blacks: Results From the Morehouse-Emory Cardiovascular (MECA) Center for Health Equity Study. *Journal of the American Heart Association*, *8*(12). doi: 10.1161/jaha.118.011633

15. Williams DR, Leavell J. (2012) The social context of cardiovascular disease: challenges and opportunities for the Jackson Heart Study. Ethn Dis. 2012;22:S1–15-21.

16. Scheier, M. F., & Carver, C. S. (1985). Optimism, coping, and health: Assessment and implications of generalized outcome expectancies. *Health Psychology*, *4*(3), 219–247. doi: 10.1037/0278-6133.4.3.219
17. Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the Life Orientation Test. *Journal of Personality and Social Psychology*, *67*(6), 1063–1078. doi: 10.1037/0022-3514.67.6.1063
18. Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology*, *57*(6), 1069–1081. doi: 10.1037/0022-3514.57.6.1069
19. Ryff, C. D., & Keyes, C. L. M. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology*, *69*(4), 719–727. doi: 10.1037/0022-3514.69.4.719
20. Campbell-Sills, L., & Stein, M. B. (2007). Psychometric analysis and refinement of the connor–davidson resilience scale (CD-RISC): Validation of a 10-item measure of resilience. *Journal of Traumatic Stress*, *20*(6), 1019–1028. doi: 10.1002/jts.20271
21. Radloff, L. S. (1977). The CES-D Scale. *Applied Psychological Measurement*, *1*(3), 385–401. doi: 10.1177/014662167700100306
22. Mujahid, M. S., Roux, A. V. D., Morenoff, J. D., & Raghunathan, T. (2007). Assessing the Measurement Properties of Neighborhood Scales: From Psychometrics to Ecometrics. *American Journal of Epidemiology*, *165*(8), 858–867. doi: 10.1093/aje/kwm040
23. Moser, A., Stuck, A. E., Silliman, R. A., Ganz, P. A., & Clough-Gorr, K. M. (2012). The eight-item modified Medical Outcomes Study Social Support Survey: psychometric evaluation showed excellent performance. *Journal of Clinical Epidemiology*, *65*(10), 1107–1116. doi: 10.1016/j.jclinepi.2012.04.007

24. K-means Cluster Analysis. (n.d.). Retrieved from https://uc-r.github.io/kmeans_clustering#gap

25. Gama, G. (n.d.). K-Medoids in R: Algorithm and Practical Examples. Retrieved from https://www.datanovia.com/en/lessons/k-medoids-in-r-algorithm-and-practical-examples/#pam-concept

26. Hierarchical Cluster Analysis. (n.d.). Retrieved from https://uc-r.github.io/hc_clustering#algorithms

27. Marina Santini (2016) Advantages & Disadvantages of k-Means and Hierarchical clustering (Unsupervised Learning). Retrieved from http://santini.se/teaching/ml/2016/Lect_10/10c_UnsupervisedMethods.pdf

28. Yiu, T. (2019, August 14). Understanding Random Forest. Retrieved from https://towardsdatascience.com/understanding-random-forest-58381e0602d2

29. Gandhi, R. (2018, July 5). Support Vector Machine - Introduction to Machine Learning Algorithms. Retrieved from https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47

30. GmbH, R. M. (n.d.). Relevance Vector Machine (RapidMiner Studio Core). Retrieved from https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/functions/relevance_vector_machine.html

31. Rozanski, A., Bavishi, C., Kubzansky, L. D., & Cohen, R. (2019). Association of Optimism With Cardiovascular Events and All-Cause Mortality. *JAMA Network Open*, *2*(9). doi: 10.1001/jamanetworkopen.2019.12200