1-1-2018

Using Cluster Analysis to Identify Subgroups of College Students at Increased Risk for Cardiovascular Disease

Dieu-My Tran

University of Nevada, Las Vegas, dieu-my.tran@unlv.edu

Kevin A. Kupzyk

University of Nebraska

Lani M. Zimmerman

University of Nebraska

Follow this and additional works at: https://digitalscholarship.unlv.edu/nursing_fac_articles

Part of the [Public Health and Community Nursing Commons](#)

Repository Citation

Tran, D., Kupzyk, K. A., Zimmerman, L. M. (2018). Using Cluster Analysis to Identify Subgroups of College Students at Increased Risk for Cardiovascular Disease. *Journal of Nursing Measurement*, 26(3), 470-482. Springer Publishing Company.

<http://dx.doi.org/10.1891/1061-3749.26.3.470>

This Article is protected by copyright and/or related rights. It has been brought to you by Digital Scholarship@UNLV with permission from the rights-holder(s). You are free to use this Article in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you need to obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/or on the work itself.

This Article has been accepted for inclusion in Nursing Faculty Publications by an authorized administrator of Digital Scholarship@UNLV. For more information, please contact digitalscholarship@unlv.edu.

Journal of Nursing Measurement

Using Cluster Analysis to Identify Subgroups of College Students at Increased Risk for Cardiovascular Disease --Manuscript Draft--

Manuscript Number:	JNM-D-17-00053R1
Full Title:	Using Cluster Analysis to Identify Subgroups of College Students at Increased Risk for Cardiovascular Disease
Article Type:	Original Study
Keywords:	cluster analysis; cardiovascular risk factors; co-occurrence; college students
Corresponding Author:	Dieu-My Thi Tran, PhD, RN, CNE University of Nebraska Medical Center Omaha, NE UNITED STATES
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	University of Nebraska Medical Center
Corresponding Author's Secondary Institution:	
First Author:	Dieu-My Thi Tran, PhD, RN, CNE
First Author Secondary Information:	
Order of Authors:	Dieu-My Thi Tran, PhD, RN, CNE
	Kevin A. Kupzyk, PhD
	Lani M. Zimmerman, PhD, RN
Order of Authors Secondary Information:	
Abstract:	<p>Background and Purpose. To examine the co-occurrence of cardiovascular risk factors and cluster subgroups of college students for cardiovascular risks. Methods. A cross-sectional, descriptive study was conducted using co-occurrence patterns and hierarchical clustering analysis in 158 college students. Results. The top co-occurring cardiovascular risk factors were overweight/obese and hypertension (10.8%, n = 17). Of the total 34 risk factors that co-occurred, 30 of them involved being overweight/obese. Six-cluster-solution was obtained, two clusters displayed elevated levels of lifetime and 30-year cardiovascular disease risks. Conclusions. The hierarchical cluster analysis identified that White, single males with a family history of heart disease, overweight/obese, hypertensive or diabetes, and occasionally (weekly) consumed red meat, take antihypertensive medication, and hyperlipidemia were considered the higher risk group compared to other subgroups.</p>

Manuscript ID JNM-D-17-00053 for Journal of Nursing Measurement Revisions/Responses
Using Cluster Analysis to Identify Subgroups of College Students at Increased Risk for Cardiovascular Disease
Revisions

Reviewer Comments	Authors' Revisions/Responses
<p>Reviewer 1:</p> <ol style="list-style-type: none"> 1. The background is quite brief. The introductory paragraph for the background should provide readers with the general bird's eye view of the entire section. Tell us what the major concepts are and then sequentially discuss them in the following paragraphs. As written, it reads more like a series of data points. Consider interpreting the data followed by a data point or two as evidence to support your conclusion. For example, I think you are saying that although we commonly think about CVD risk among middle aged to older adults, it is problematic among young adults such as college students. In general, the background also could be improved in terms of flow of ideas if the authors crafted topic sentences that introduced the topic of each paragraph. Beginning paragraphs with sentences like "Kestila et al examined the associations..." does not adequately introduce the topic of risk at a young age (I think this is the direction of the paragraph). 2. Page 4, line 46. How was reliability of the cluster analysis determined in the study by Shah? 3. Methods: What is the rationale for excluding students who weighed less than 110 pounds? 	<p>Reviewer 1:</p> <ol style="list-style-type: none"> 1. Thank you for your insights and suggestions. The background section has been revised in red to strengthen the flow and organization of this section. 2. Unfortunately, the reliability of the cluster analysis was not mentioned in the study done by Shah et al. Cluster analysis is a statistical technique; therefore, reliability most likely will be not provided. 3. The rationale of excluding students who weigh less than 110 pounds is because of the blood drawn. Subjects have to be at least 110 pounds to complete the blood draw based on our IRB requirements.

<p>4. With regard to the survey questions, were these open-ended (fill in the blank) or did you have them rate their physical activity in some way? There are several valid and reliable indexes that could be used to quantify physical activity.</p> <p>5. BMI was classified ranging from overweight to several classes of obesity. What about the normal or below normal BMI students?</p> <p>6. What was the rationale for not standardizing fasting or non fasting glucose levels? I am guessing that you dichotomized the glucose into elevated or not. Was this your definition of diabetes in the analysis? Or was the diabetes variable by health history?</p> <p>7. Results: Your purpose was to identify modifiable and non-modifiable risk factors, but these are not discussed in the results or the discussion.</p> <p>8. Discussion: Consider beginning your discussion with your most compelling findings and then compare and contrast with published evidence. Do your results really differ from other age groups? Do you think they differ when comparing college students to similarly aged people who do not go to college?</p> <p>9. Tables and figures: Table 2. What does the mean of physical activity mean? Hours per week? Days? Did you differentiate between low, moderate and high intensity?</p>	<p>4. Our physical activity questionnaire was brief, please refer to the Study Variables section for details on the physical activity questionnaire. This is one of the weaknesses of our study for not including a published validated physical activity tool. Instead we created our own brief physical activity questionnaire.</p> <p>5. Please refer to our addition in red under the Methods and Results sections to include the normal weight and underweight BMI.</p> <p>6. We standardized our glucose levels because we obtained fasting or non-fasting glucose levels from subjects depending on their fasting status during the blood draw. We did not require subjects to be fasting to be included in our study. We did dichotomize based on the glucose level that was drawn into elevated or not. This was mentioned under the Methods, Anthropometric and Cardiovascular Risk Factor Assessments section.</p> <p>7. To clarify, the purpose of our study is to examine the co-occurrence of cardiovascular risk factors among college students, ages 19 to 39, and to cluster subgroups of college students for cardiovascular risks based on socio-demographics, non-modifiable and modifiable risk factors. We aim to examine the co-occurrence and cluster the subgroups using the variables non-modifiable and modifiable risk factors.</p> <p>8. We have included a Nursing Implications section under the Discussion to strengthen this section.</p> <p>9. The mean of physical activity is measured in how often per week did they complete some type of physical activity. We also asked the subjects what level of physical activity they performed with the following choices: sedentary, low level (daily walking), moderate level activity (walking or jogging for 30 minutes, and vigorous level activity (jogging for more than 30 minutes).</p>
---	--

10. The dendrogram may be difficult to interpret for some readers. I have seen some dendrograms color coded to help readers see the clusters. Likely color coding will not be an option, but may you could shade or label the 6 clusters to assist readers unfamiliar with this type of figure.

11. Mechanics: The AHA statistics citation is incorrectly reported. There is a lead author that you can cite followed by the co-authors (as opposed to "Writing group members..."). There is a 2017 version of this document that should be updated.

12. The citation for the AHA 2013 CV risk assessment is incorrect. The lead author is Goff, D. and the journal and accompanying data are missing from the citation as well.

Reviewer 2:

1. There are several suggestions for the authors, which would make it easier for the reader to understand the research with the following responses –
Background: Literature review contained relevant background information. However, the first paragraph on page 3 was not clear. It was very confusing.
2. Results: Need to give more information on the demographic. HTN and overweight/obese are the two highest risk factors of cardiovascular disease not only for college students, but also for young and old adults. In addition, although there is a lack of study in college students for risk factors of cardiovascular problems, the study findings are not surprising.

10. We explored a few options based on the given suggestions and we modified the dendrogram with a reference line indicating where the 6 clusters are separated.

11. The following citation has been corrected in red and updated to the 2017.

12. Thank you for bringing this to our attention. The following reference has been corrected in red.

Reviewer 2:

1. The first paragraph on page 3 beginning with, "Several research methods are available to identify or predict individuals at increased risk individuals of chronic illnesses" has been revised in red for clarity.

2. Based on our study demographic data that was collected, it was reported under the Sample section. The second highest race/ethnicity was added in red under the Sample section.

Thank you for your insight. We completely agree with you. Even though the findings of high prevalence of HTN and overweight/obese is not surprising, the lack of attention in this population is one of the reasons our research team aim to accomplish by conducting research in college students.

Comments from the Editor:

This is the Journal of NURSING Measurement yet nowhere in the manuscript is nursing even mentioned! If you chose to revise this please make the nursing implications for research, practice and education clear to the reader, as appropriate throughout the paper.

Comments from the Editor:

We apologize for the oversight. We have added a Nursing Implications section to address this concern.

Title Page for Journal of Nursing Measurement

Manuscript Title: Using Cluster Analysis to Identify Subgroups of College Students at Increased Risk for Cardiovascular Disease

Running head: Using Cluster Analysis

Dieu-My T. Tran, ¹ Kevin A. Kupzyk, ² and Lani M. Zimmerman ³

¹ University of Nevada, Las Vegas – School of Nursing, Las Vegas, Nevada, United States PhD, RN, CNE email: dieu-my.tran@unlv.edu

² University of Nebraska Medical Center – College of Nursing, Omaha, Nebraska, United States PhD email: kevin.kupzyk@unmc.edu

³ University of Nebraska Medical Center – College of Nursing, Lincoln, Nebraska, United States PhD, RN, FAAN email: lzimmerm@unmc.edu

Corresponding Author:

Dieu-My T. Tran, PhD, RN, CNE
University of Nevada, Las Vegas
4505 S. Maryland Parkway Box 453018
Las Vegas, NV 89154-3018
Dieu-my.tran@unlv.edu
Phone: (702) 895-3371
Fax: (702) 895-4807

Acknowledgements – The authors extend appreciation to Scott Shurmur, MD, Carol Pullen, EdD, and Bernice Yates, PhD, as supervisory committee members at the University of Medical Center – College of Nursing, for their guidance and mentorship with the study.

Conflict of Interest statement - No conflict of interest has been declared by the authors.

Number of words in text: 3674

2 Tables, 3 Figures

Using Cluster Analysis to Identify Subgroups of College Students at Increased Risk for Cardiovascular Disease

Cardiovascular disease (CVD) accounts for more deaths than any other leading cause of death in the United States (Benjamin et al., 2017). On average, approximately 2,200 Americans die of CVD each day with an average of one death every 40 seconds. An estimated 80% of CVDs are preventable by reducing or eliminating cardiovascular (CV) risk factors (i.e. smoking cessation). Eating a healthy diet, engaging in physical activity, and maintaining a healthy weight in addition to controlling high blood pressure, high glucose levels, and elevated lipid levels can aid in reducing and eliminating CV risk factors (Benjamin et al., 2017).

According to American College Health Association National College Health Assessment (2016) report on 95,761 college students, 22.9% were overweight and 13.9% were obese. Additionally, 1.2% of college students reported being diagnosed with or treated for diabetes, 3.2% diagnosed with or treated for high blood pressure, and 2.9% diagnosed with or treated for high cholesterol. Also, more than 50% of the college students did not meet the recommended physical activity guidelines. According to the Coronary Artery Risk Development in Young Adults (CARDIA) study, young adults ($n = 3,014$) who were overweight or obese had a lower health-related quality of life at the 20-year follow-up compared to normal-weight subjects (Kozak et al., 2011). Furthermore, the transition from home to the college environment, along with different cultural and social situations, can potentially influence lifestyle choices of college students when it comes to diet, alcohol consumption, physical activity, which can have life-long consequences on their CV health (Collins, Dantico, Shearer, & Mossman, 2004).

The purpose of this study was to examine the co-occurrence of CV risk factors among college students, aged 19 to 39 years, and to cluster subgroups these students for CV risks based

on socio-demographics, and non-modifiable and modifiable risk factors. The overall goal was to identify a target group of individuals at an increased risk for CVD.

BACKGROUND

CVD is one of the most common chronic diseases among middle aged to older adults. However, it is also problematic for young adults including college students. Kestila et al. (2012) examined the associations between socioeconomic status and risk factors for reduced CV risk in young adults in Finland (aged 24-39 years) at baseline and six-year follow-up. They found an association between higher educational level in young adults and favorable CV risk factors at baseline and six-year follow-up. In turn, a lower educational level would suggest that identifying CV risk factors in an isolated college population would translate into more issues for the general young adult population. According to the 2017 Heart Disease and Stroke statistics report, modifying CVD risk factors at a young age and evaluating potential benefits of intensive prevention programs for young and middle-aged adults at risk, can potentially increase the likelihood of health longevity (Benjamin et al., 2017). In turn, college students are an important population to examine to identify potential CV risk factors and prevention strategies.

Promoting and practicing healthy behaviors especially among young adults, is one of the most effective methods to reduce chronic illnesses such as CVD, particularly focusing on the risk factors (Ulla Diez, Fortis, & Franco, 2012). According to Daviglus et al., (2004), women aged 18 to 39 years with optimal levels for all five major CV risk factors (blood pressure, cholesterol level, body mass index [BMI], diabetes mellitus, and smoking status) are at lower risk of developing coronary heart disease (CHD) and CVD. However, most CV risk factors tend to co-occur (Zubair, Kuzawa, Lee, McDade, & Adair, 2014). Limited studies have investigated which CV risk factors co-occur or cluster together in the young adult population. Thus far, Zubair et al.

(2014) used a K-means clustering technique to examine the cardiometabolic profile of 1,621 Filipino young adults. They reported cluster patterns of cardiometabolic risk factors. Specifically, adiposity had a strong association with a cluster characterized by insulin resistance and high triglyceride levels. Furthermore, overweight individuals were more likely to have high blood pressure (Zubair et al., 2014).

Several research methods are available to identify or predict individuals at increased risk individuals of chronic illnesses. One in particular, is a cluster analysis. A cluster analysis is a technique that identifies subgroups with similar characteristics within a larger group based on the variables entered into the analysis (Cabell, Justice, Konold, & McGinty, 2011). This type of analysis is an important method in cancer and elderly cardiac patients research to understand the impact of symptoms and quality of life (Fox & Lyon, 2006; Fukuoka, Lindgren, Rankin, Cooper, & Carroll, 2007; Miaskowski et al., 2017; Miaskowski et al., 2006). However, this method is seldom utilized in CV risk factors research. Likewise, cluster analysis can be used to identify subgroups with higher CV risk profiles that need more attention and intervention strategies to prevent CVD. Shah et al. (2011) examined CV risk factors in adolescents and young adults aged 11 to 23 years and utilized cluster analysis to cluster CV risks into low- and high-risk groups. They found that the high-risk group had increased vascular thickness and stiffness compared to the low-risk group. As variables for clustering, Shah et al. (2011) included BMI, lipid profile, blood pressure, glucose, and insulin; this study found that cluster analysis was a reliable tool to assess abnormal vascular function. Furthermore, the study showed that clustering CV risks was a useful method to predict atherosclerosis and CHD (Shah et al., 2011). Therefore, clustering analysis is potentially an innovative way to identify and locate higher risk groups to target effective risk-reduction interventions.

METHODS

This was a prospective cross-sectional study which enrolled a convenience sample of 158 college students (aged 19-39 years) from a large Midwestern university campus. College students were recruited via research flyers, recommendations from the health care staff at the university, and word-of-mouth. Students were enrolled between July and August 2014. Exclusion criteria were students who were diagnosed with CVD such as myocardial infarction, stroke, or CHD; pregnant or weighed less than 110 lbs. The study was approved by the Institutional Review Board of the university and all subjects provided informed consent. This is a continuation of data analysis from a previously published manuscript (Tran et al., 2016).

Study Variables

Socio-demographics and Health History

We collected demographic characteristics from all subjects (age, sex, race/ethnicity, marital status, and insurance status) as well as their personal health history (past medical health history, family history of heart disease, antihypertensive medication to control high blood pressure, and smoking status). We also asked the subjects five questions to capture dietary habits (three questions) and physical activity (two questions) followed by four responses: “What is your level of physical activity?” “If you do some type of physical activity, how often?” “How often do you eat out per week?” “How often do you drink sugary beverages per week?” and “How often do you eat red meats per week?”

Anthropometric and Cardiovascular Risk Factor Assessments

A stadiometer and digital scale were used to measure height (inches) and weight (lbs). An online calculator by the National Institutes of Health was used to convert height and weight into BMI (US Department of Health and Human Services, National Institutes of Health, n.d.). BMI

was classified as overweight (25 to 29.9 kg/m²), obese class I (30 to 34.9 kg/m²), obese class II (35 to 39.9 kg/m²), and obese class III (≥ 40 kg/m²; Go et al., 2013). Blood pressure was measured according to the guidelines by the American Heart Association using a mercury sphygmomanometer (Weber et al., 2014). Subjects were seated and arms were supported at the level of the heart, and blood pressure was measured before the venous puncture procedure to minimize levels of pain and stress. Blood pressure was classified as normal if systolic blood pressure (SBP) ≤ 120 mm Hg or diastolic blood pressure (DBP) ≤ 80 mm Hg, prehypertension if SBP between 120-139 mm Hg or DBP between 80 to 89 mm Hg, or hypertensive if SBP ≥ 140 mm Hg or DBP ≥ 90 mm Hg (Weber et al., 2014).

After completion of questionnaires, we obtained 10 mL of venous blood from fasting and non-fasting subjects for the measurement of lipid profiles and glucose levels. We also collected venous blood into tubes containing ethylenediaminetetraacetic acid and stored under refrigerated conditions. The university health center's laboratory technician conducted lipid profiles and glucose analyses. The lipid profile included total cholesterol, low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), and triglycerides. We classified normal and elevated levels according to the National Cholesterol Education Program Adult Treatment Panel III (2002). We classified blood glucose according to the Diagnosis and Classification of Diabetes Mellitus (2016). We classified fasting and non-fasting glucose levels as elevated if values were ≥ 106 mg/dL and ≥ 200 mg/dL, respectively (ADA, 2010). Approximately, one-third of the subjects' blood samples were fasting.

Risk Assessments

In this study, two risk assessments were measured to estimate CVD risk using the American College of Cardiology/American Heart Association Atherosclerotic Cardiovascular

Disease (ASCVD) Score (Goff et al., 2014) and the 30-year CVD risk assessment (Pencina, D'Agostino, Larson, Massaro, & Vasan, 2009). Both assessments used the following variables: sex (male vs. female), age (years), race (White and other races vs. African American), total cholesterol (mg/dL), HDL-C (mg/dL), SBP (mm Hg), antihypertensive medications (yes vs. no), diabetes (yes vs. no), and smoker (yes vs. no; Goff et al., 2014). Race was not included in the 30-year risk assessment (Pencina et al., 2009). The lipids-dependent equations were used from the 30-year CV risk assessment to calculate each individual risk in 30 years. There were two risks generated, hard and full CVD risk (Pencina et al., 2009). The hard CVD is defined as having a myocardial infarction, coronary death, and stroke whereas the full CVD covered the hard CVD or coronary insufficiency, transient ischemic attack, angina pectoris, congestive heart failure, or intermittent claudication (Pencina et al., 2009). The ASCVD score provides an estimated 10-year and lifetime risk for ASCVD, defined as coronary death or non-fatal myocardial infarction, or fatal or non-fatal stroke (Stone et al., 2014).

Statistical Analysis

Data were analyzed using IBM SPSS version 20 software. Statistical significance was determined if the p-value was less than .05. Co-occurrence of CV risk factors was investigated using frequencies of each unique pattern of risk factors.

Cluster Analysis

Hierarchical clustering analysis was used to identify the subgroups for similarity and dissimilarity of dichotomous and ordinal variables (Frades & Matthiesen, 2010; Fukuoka et al., 2007). Hierarchical clustering was the selected approach because there was insufficient literature to suggest the number of clusters of college students to the group regarding CV risk factors (Devlin, McNulty, Nugent, & Gibney, 2012). The agglomeration schedule and dendrogram

(visual representation generated by possible groups in various levels) were examined (Frades & Matthiesen, 2010). Next, Ward's method was used to cluster cases based on a Euclidean distance matrix. The variable transformation involved was a z-score standardization to obtain common scale units across variables. The appropriate number of clusters was determined based on the dendrogram and interpretability of the cluster solution (Devlin et al., 2012).

We included the following variables in the clustering analysis: sex, race/ethnicity (White vs. non-White), insurance status, past medical history of heart problems (other than CVD; e.g. heart murmurs), family history of heart disease, antihypertension medication, diabetic, hypertension, hyperlipidemia, current smoker, overweight/obese, marital status, physical activity levels, how often subjects eat out, eat red meat, and drink sugary beverages per week. These variables were selected because they were dichotomous variables that included the sample socio-demographic characteristics (including health history) and non-modifiable and modifiable CV risk factors. Analysis of Variance (ANOVA) and Chi-square tests were used to interpret the subgroup profiles by utilizing the lifetime ASCVD risk and 30-year CVD risk estimates as the dependent variables.

RESULTS

Sample

Of those who were recruited, a total of 169 college students responded, two students did not meet the inclusion criteria, and nine students withdrew from the study. A total of 158 college students, aged 19 to 39 years ($M = 24.3$, $SD = 4.6$) were in the final sample. More than half of the subjects were male (54.4%, $n = 86$) and White (63.1%, $n = 99$) followed by Asian/Pacific Islander (13.4%, $n = 21$). The vast majority of the subjects were single (82.9%, $n = 131$) and had insurance coverage (93.0%, $n = 146$). Approximately 32.3% ($n = 51$) of the subjects reported

having a family history of heart disease. The average 30-year hard CVD risk assessment was 2.3%, 30-year full CVD risk assessment was 4.8%, and the lifetime risk estimate was 31.4%.

Co-occurrence of Risk Factors

With respect to the five standard CV risk factors measured, the most frequent risk factors were being overweight/obese (44.9%, $n = 71$ [**obese $n = 26$; overweight $n = 45$; normal weight $n = 84$; and underweight $n = 3$**]), followed by high blood pressure (12.7%, $n = 20$), current smoker (7.0%, $n = 11$), high glucose level (3.8%, $n = 6$), and elevated lipid levels (3.2%, $n = 5$). When the five standard CV risk factors were profiled together, 75 subjects (47.5%) had no CV risk factors, 57 subjects (36.1%) had at least one risk factor, 22 subjects (13.9%) had two risk factors, and four subjects (2.5%) had three risk factors.

The study sample showed a total of 34 risk factor co-occurrences. The top two co-occurring CV risk factors were overweight/obese and high blood pressure (10.8%, $n = 17$), followed by overweight/obese and current smokers (3.8%, $n = 6$), and then overweight/obese and high glucose level (2.5%, $n = 4$). As shown in Table 1, of the total 34 risk factors that co-occurred, 30 of them involved being overweight/obese.

Cluster Analysis Subgroups

Based on the generated dendrogram, a six-cluster-solution was retained. While there is no statistical test for determining the optimal number of clusters, examining the dendrogram from top to bottom, the full sample is repeatedly split into smaller subgroups. The distance in between splits is interpreted as the degree of similarity within a cluster. The longer it takes to split a cluster, the more cohesive the cluster is, regarding similarity on the measured variables. Our analysis found six subgroups that had formed near the top of the dendrogram with a considerable amount of distance before the groups continued to split further (Figure 1). These six subgroups

were retained for further analysis and interpretation. The group sizes ranged from three to 65, with 3 of the subgroups having less than 10 members. Two subgroups accounted for over 80% of the sample.

Table 2 contains the cluster analysis subgroups, of the six-cluster-solution, three cluster subgroups were important for further discussion due to elevated risk index scores and high frequency of risk factors (i.e., clusters 2, 3 and 6). Cluster 2 included all the subjects in the study taking antihypertensive medication ($n = 5$). Cluster 3 ($n = 60$) included White, single males with a family history of heart disease, overweight/obesity, having high blood pressure or high glucose levels, and a diet of red meat more than three times per week. Cluster 6 included all the subjects in the study demonstrating elevated lipid levels ($n = 5$). As shown in figures 2 and 3, the results from the ANOVA means plot for the lifetime ASCVD and 30-year CVD risk assessment, clusters 2 and 6 were the groups with the highest risk. Still, the sample in both clusters had less than 10 subjects. Based on the risk profile, cluster 3 ($n = 60$) was considered the subgroup with more risk factors to target for CV risk reduction intervention because the cluster consisted of White, single males with a family history of heart disease, overweight/obese, high blood pressure or high glucose levels, and ate red meat weekly.

DISCUSSION

There are limited studies examining CV risk factors using clustering analysis, and virtually none in college students. This study brings forth well-known techniques not yet applied to this particular population, which could benefit practice and research. Cluster analysis has been applied extensively in nutrition studies to derive dietary patterns (Devlin et al., 2012; Ocké, 2013; Quatromoni et al., 2002), and cancer symptoms and quality of life (Fox & Lyon, 2006; Fukuoka et al., 2007; Miaskowski et al., 2006). Most commonly, cluster analysis has been used

as a patterning method in the healthy population (Devlin et al., 2012). As well, cluster analysis results in mutually exclusive non-overlapping clusters of individuals (Devlin et al., 2012; Ocké, 2013). Applying cluster analysis to identify subgroups who are at high risk of CVD events in our study also has significant implications. This study found that college students who were taking antihypertensive medications, had elevated lipid levels in addition to being White, single males with a family history of heart disease, overweight/obese, high blood pressure or high glucose levels, and ate red meat weekly, had similar characteristics that put them at an increased risk for CVD. Despite the fact that the ANOVA means plot only identified clusters 2 and 6 as having increased risks, placing sample size into consideration, cluster 3 would also be a strong subgroup to aim for intervention programs to reduce CVD mortality and morbidity. Further research utilizing the clustering technique is warranted in the college population which may decrease the CVD burden when interventions are implemented appropriately.

More than 50% of the subjects had one or more CV risk factors; the most commonly occurring CV risk factors were overweight/obesity and high blood pressure. Further, to add to the epidemic health issue of obesity, of the 34 CV risk factors that co-occurred, 30 were related to being overweight/obese. Obesity continues to be a problem throughout life but often starts during younger ages (Lytle et al., 2014). Thus, overweight young adults are exposed to long-term CV risks (Berger, Jordan, Lloyd-Jones, & Blumenthal, 2010). Liu et al. (2012) studied the CARDIA subjects to examine healthy lifestyle factors at baseline, seven-year follow-up, and 20-year follow-up, and reported that, regardless of race or sex, individuals who adopted and maintained a healthy lifestyle with no CV risk factors during young adulthood had a reduced CV risk profile during middle-age years. Likewise, Liu et al. (2012) reported subjects with increased CV risk factors during young adulthood presented a significantly increased prevalence of CV

risk profile during middle-age years. The co-occurrences of being overweight/obese and having high blood pressure are consistent with Liu et al., (2012) showing an association between obesity and high blood pressure.

The two most commonly co-occurring CV risk factors with overweight/obese are being a current smoker and having high blood pressure. Similarly, Matsha, Hassan, Kidd, and Erasmus (2012) reported that smoking was more prevalent in young adults between ages 20 to 30 years; the prevalence of being overweight was similar across age groups (aged 20-60 years). Interestingly, Marma et al. (2010) analyzed the National Health and Nutrition Examination Survey (NHANES) data using the 10-year Framingham Risk score for stratification and reported that no single CV risk factor in young adults was responsible for predicting lifetime CV risks. Gooding and de Ferranti (2010) examined young adults aged 18 to 30 years ($n = 2164$) from the CARDIA study sample and reported the majority of young adults with optimal CV health profile lost more than one ideal healthy lifestyle by middle age, with higher probability if the young adults were overweight/obese and were smokers during the baseline assessment. In essence, maintaining a low-risk CV profile during young adulthood is thought to be ideal for preventing and delaying atherosclerosis and CVD during middle age and older years (Loria et al., 2007).

Nursing Implications

We found that college students are currently at risk for CVD and the most common co-occurring risks are being overweight/obese and having high blood pressure. These findings are not surprising as CVD remains the number one cause of deaths with hypertension being the most prevalent CVD risk factor (Benjamin et al., 2017; Jain et al., 2014). It may not be surprising to discover these findings; however, it is surprising that little is being done for the young adult

college students. This age group is where modifiable risks develop and if not avoided will develop into chronic diseases in middle to older adulthood. Our study highlights the need to study young adult college students at risk for CVD, with the implication that different subgroups may have increased CV risk. Nurses working in healthcare centers of universities or nurses performing screenings at health fairs have the potential to identify at risk college students for CV risk factors. This study represents an initial step in the direction of focusing on CVD in young adulthood with the implications that nurses could be at the forefront to address this problem.

Limitations

This was a cross-sectional, descriptive study; therefore, generalizability as a whole is limited. Another limitation to consider is the lack of statistical tests to determine the correct number of cluster solutions. However, an increasing number of studies utilizing this method in college students with CV risk factors may support the appropriate cluster solutions. Additional studies in this area needs to be done to determine the consistency of the subgroups.

Conclusions

Detecting high risk groups through a clustering technique has the potential to identify groups of college students to target for interventions. Based on the co-occurrences of CV risk factors and the cluster solution in the current study, White, single males who are overweight/obese, have high blood pressure, have high glucose levels, a family history of heart disease along with eating red meat more than two times a week, take antihypertensive medication, have elevated lipid levels were identified as the highest risk group to target for CV risk reduction intervention. Additionally, overweight was the most prevalent CV risk factor in this population, which should not be ignored. Utilizing cluster analysis and co-occurrence of CV risk factors are valuable statistical methods to identify subgroups who are at increased CVD

risks. To date, very few studies have used this method. Clustering technique such as hierarchical clustering analysis can be useful in identifying high-risk CVD groups for targeting interventions in college students.

References

ADA. (2010). Diagnosis and classification of diabetes mellitus. *Diabetes Care*, 33(1S), S62-9S.

American College Health Association. (2016). *American college health association-national college health assessment II: Reference group executive summary spring 2016*. Hanover, MD: American College Health Association.

American Diabetes Association. (2016). Strategies for improving care. sec. 1. in standards of medical care in diabetes-2016. *Diabetes Care*, 39(Suppl.1), S6-S12. doi: 10.2337/dc16-S001

Benjamin, E. J., Blaha, M. J., Chiuve, S. E., Cushman, M., Das, S. R., Deo, R., . . . American Heart Association Statistics Committee and Stroke Statistics Subcommittee. (2017). Heart disease and stroke statistics-2017 update: A report from the American Heart Association. *Circulation*, 135(10), e146-e603. doi:10.1161/CIR.0000000000000485

Berger, J. S., Jordan, C. O., Lloyd-Jones, D., & Blumenthal, R. S. (2010). Screening for cardiovascular risk in asymptomatic patients. *Journal of the American College of Cardiology*, 55(12), 1169-1177.

Cabell, S. Q., Justice, L. M., Konold, T. R., & McGinty, A. S. (2011). Profiles of emergent literacy skills among preschool children who are at risk for academic difficulties. *Early Childhood Research Quarterly*, 26(1), 1-14.

Collins, K. M., Dantico, M., Shearer, N. B. C., & Mossman, K. L. (2004). Heart disease awareness among college students. *Journal of Community Health*, 29(5), 405-420.

Daviglus, M. L., Stamler, J., Pirzada, A., Yan, L. L., Garside, D. B., Liu, K., . . . Greenland, P. (2004). Favorable cardiovascular risk profile in young women and long-term risk of cardiovascular and all-cause mortality. *Jama*, 292(13), 1588-1592.

Devlin, U. M., McNulty, B. A., Nugent, A. P., & Gibney, M. J. (2012). The use of cluster analysis to derive dietary patterns: Methodological considerations, reproducibility, validity and the effect of energy mis-reporting. *Proceedings of the Nutrition Society*, 71(4), 599.

Fox, S. W., & Lyon, D. E. (2006). Symptom clusters and quality of life in survivors of lung cancer. *Oncology Nursing Forum*, 33(5) 931-936.

Frades, I., & Matthiesen, R. (2010). Overview on techniques in cluster analysis. *Bioinformatics methods in clinical research* (pp. 81-107) Springer.

Fukuoka, Y., Lindgren, T. G., Rankin, S. H., Cooper, B. A., & Carroll, D. L. (2007). Cluster analysis: A useful technique to identify elderly cardiac patients at risk for poor quality of life. *Quality of Life Research*, 16(10), 1655-1663.

Go, A. S., Mozaffarian, D., Roger, V., Benjamin, E. J., Berry, J. D., Borden, W. B., . . . Turner, M. B. (2013). Heart disease and stroke statistics--2013 update: A report from the American Heart Association. *Circulation*, 127(1), e6-e245. doi:10.1161/CIR.0b013e31828124ad

Goff, D. C., Jr, Lloyd-Jones, D. M., Bennett, G., Coady, S., D'Agostino, R. B., Gibbons, R., . . . American College of Cardiology/American Heart Association Task Force on Practice Guidelines. (2014). 2013 ACC/AHA guideline on the assessment of cardiovascular risk: A report of the American College of Cardiology/American Heart Association task force on

practice guidelines. *Circulation*, 129(25 Suppl 2), S49-73.

doi:10.1161/01.cir.0000437741.48606.98 [doi]

Gooding, H. C., & de Ferranti, S. (2010). Cardiovascular risk assessment and cholesterol management in adolescents: Getting to the heart of the matter. *Current Opinion in Pediatrics*, 22(4), 398-404. doi:10.1097/MOP.0b013e32833a6e22

Jain, J., Mathur, H., Apte, S., Sinha, U., Agarwal, K., & Chanchlani, R. (2014). Prevalence of prehypertension and hypertension and its correlation with anthropometric measurements in medical students of central India. *Journal of Evolution of Medical and Dental Sciences*, 3(10), 2429-2433.

Kestilä, P., Magnussen, C. G., Viikari, J. S. A., Kähönen, M., Hutri-Kähönen, N., Taittonen, L., . . . Raitakari, O. T. (2012). Socioeconomic status, cardiovascular risk factors, and subclinical atherosclerosis in young adults: The cardiovascular risk in young Finns study. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 32(3), 815-821.
doi:10.1161/ATVBAHA.111.241182

Kozak, A. T., Daviglus, M. L., Chan, C., Kiefe, C. I., Jacobs, D., & Liu, K. (2011). Relationship of body mass index in young adulthood and health-related quality of life two decades later: The coronary artery risk development in young adults study. *International Journal of Obesity*, 35(1), 134-141.

Liu, K., Daviglus, M. L., Loria, C. M., Colangelo, L. A., Spring, B., Moller, A. C., & Lloyd-Jones, D. M. (2012). Healthy lifestyle through young adulthood and the presence of low cardiovascular disease risk profile in middle age: The coronary artery risk development in

(young) adults (CARDIA) study. *Circulation*, 125(8), 996-1004.

doi:10.1161/CIRCULATIONAHA.111.060681

Loria, C. M., Liu, K., Lewis, C. E., Hulley, S. B., Sidney, S., Schreiner, P. J., . . . Detrano, R.

(2007). Early adult risk factor levels and subsequent coronary artery calcification: The CARDIA study. *Journal of the American College of Cardiology*, 49(20), 2013-2020.

Lytle, L. A., Moe, S. G., Nanney, M. S., Laska, M. N., Linde, J. A., Petrich, C. A., & Sevcik, S.

M. (2014). Designing a weight gain prevention trial for young adults: The CHOICES study. *American Journal of Health Education*, 45(2), 67-75.

Marma, A. K., Berry, J. D., Ning, H., Persell, S. D., & Lloyd-Jones, D. (2010). Distribution of 10-year and lifetime predicted risks for cardiovascular disease in US adults: Findings from the national health and nutrition examination survey 2003 to 2006. *Circulation*.

Cardiovascular Quality and Outcomes, 3(1), 8-14.

doi:10.1161/CIRCOUTCOMES.109.869727

Matsha, T. E., Hassan, M. S., Kidd, M., & Erasmus, R. T. (2012). The 30-year cardiovascular risk profile of South Africans with diagnosed diabetes, undiagnosed diabetes, pre-diabetes or normoglycaemia: The Bellville, South Africa pilot study: Cardiovascular topics.

Cardiovascular Journal of Africa, 23(1), 5-11.

Miaskowski, C., Barsevick, A., Berger, A., Casagrande, R., Grady, P. A., Jacobsen, P., . . .

Marden, S. (2017). Advancing symptom science through symptom cluster research: Expert panel proceedings and recommendations. *Journal of the National Cancer Institute*, 109(4), 10.1093/jnci/djw253. Print 2017 Apr. doi: djw253 [pii]

- Miaskowski, C., Cooper, B. A., Paul, S. M., Dodd, M., Lee, K., Aouizerat, B. E., . . . Bank, A. (2006). Subgroups of patients with cancer with different symptom experiences and quality-of-life outcomes: A cluster analysis. *Oncology Nursing Forum*, 33(5), E79-E89.
- National Cholesterol Education Program (NCEP) Expert Panel (Adult Treatment Panel III). (2002). Third report of the national cholesterol education program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (adult treatment panel III): Final report. *Circulation*, 106(25), 3143-3421.
- Ocké, M. C. (2013). Evaluation of methodologies for assessing the overall diet: Dietary quality scores and dietary pattern analysis. *Proceedings of the Nutrition Society*, 72(02), 191-199.
- Pencina, M. J., D'Agostino, R. B., Larson, M. G., Massaro, J. M., & Vasan, R. S. (2009). Predicting the 30-year risk of cardiovascular disease the Framingham Heart Study. *Circulation*, 119(24), 3078-3084.
- Quatromoni, P. A., Copenhafer, D. L., Demissie, S., D'Agostino, R. B., O'Horo, C. E., Nam, B., & Millen, B. E. (2002). The internal validity of a dietary pattern analysis. the Framingham nutrition studies. *Journal of Epidemiology and Community Health*, 56(5), 381.
- Shah, A. S., Dolan, L. M., Gao, Z., Kimball, T. R., & Urbina, E. M. (2011). Clustering of risk factors: A simple method of detecting cardiovascular disease in youth. *Pediatrics*, 127(2), e312-e318. doi:10.1542/peds.2010-1125
- Stone, N. J., Robinson, J. G., Lichtenstein, A. H., Bairey Merz, C. N., Blum, C. B., Eckel, R. H., . . . American College of Cardiology/American Heart Association Task Force on Practice

Guidelines. (2014). 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: A report of the American College of Cardiology/American Heart Association task force on practice guidelines. *Circulation*, 129(25 Suppl 2), S1-45. doi:10.1161/01.cir.0000437738.63853.7a [doi]

Tran, D. T., Zimmerman, L. M., Kupzyk, K. A., Shurmur, S. W., Pullen, C. H., & Yates, B. C. (2016). Cardiovascular risk factors among college students: Knowledge, perception, and risk assessment. *Journal of American College Health*, 65(3), 158-167. doi: 10.1080/07448481.2016.1266638

Ulla Diez, S. M., Fortis, A. P., & Franco, S. F. (2012). Efficacy of a health-promotion intervention for college students: A randomized controlled trial. *Nursing Research*, 61(2), 121-132. doi:10.1097/NNR.0b013e3182475aaa [doi]

U.S. Department of Health and Human Services, National Institutes of Health. (n.d.) Calculate your body mass index. Retrieved from https://www.nhlbi.nih.gov/health/educational/lose_wt/BMI/bmicalc.htm

Weber, M. A., Schiffrin, E. L., White, W. B., Mann, S., Lindholm, L. H., Kenerson, J. G., . . . Ram, C. V. S. (2014). Clinical practice guidelines for the management of hypertension in the community. *The Journal of Clinical Hypertension*, 16(1), 14-26.

Zubair, N., Kuzawa, C. W., Lee, N. R., McDade, T. W., & Adair, L. S. (2014). Clustering and determinants of cardiometabolic risk factors among filipino young adults. *Asia Pacific Journal of Clinical Nutrition*, 23(1), 148.

Using Cluster Analysis to Identify Subgroups of College Students at Increased Risk for Cardiovascular Disease

Tables and Figures

Table 1. *Co-occurrence of Risk Factors*

	<i>n</i>	% of sample
Overweight & High blood pressure	17	10.76%
Overweight & Smoking	6	3.80%
Overweight & High glucose levels	4	2.53%
Overweight & Elevated lipid levels	3	1.90%
High glucose levels & Blood pressure	2	1.27%
High blood pressure & Smoking	1	0.63%
High glucose levels & Elevated lipid levels	1	0.63%

Note: $n = 26$ had two or more risk factors.

Table 2. Cluster Analysis Subgroups

	Cluster 1 <i>n</i> =65	Cluster 2 <i>n</i> =5	Cluster 3 <i>n</i> =60	Cluster 4 <i>n</i> =3	Cluster 5 <i>n</i> =16	Cluster 6 <i>n</i> =5	p-value
Gender							<.001
Male	19	5	49	2	7	3	
Female	46	0	11	1	9	2	
Race/Ethnicity							<.001
White	29	4	50	2	9	2	
Non-White	36	1	10	1	7	3	
Marital status							.034
Married/Living together	18	1	4	0	2	0	
Single/Divorced	47	4	56	3	14	5	
Insurance (Yes)	65	5	60	3	6	4	<.001
PMH - Heart Problems (Yes)	0	0	0	3	0	0	<.001
Family Hx of Heart Disease (Yes)	19	2	22	1	3	3	0.546
Taking Antihypertensive (Yes)	0	5	0	0	0	0	<.001
High glucose levels (Yes)	0	0	4	0	0	1	0.085
Overweight/Obese (Yes)	8	4	43	2	9	3	<.001
High blood pressure (Yes)	0	3	15	1	1	0	<.001
Elevated lipid levels (Yes)	0	0	0	0	0	5	<.001
Current Smoker (Yes)	0	0	1	0	10	0	<.001
Physical Activity		<i>M</i> =3.6		<i>M</i> =3.33			.037
Eat Out							.047
Drink Sugary Beverages							.154
Eat Red Meat			<i>M</i> =2.9			<i>M</i> =2.4	<.001

Note: PMH Past medical history; Hx history

Figure 1. *Dendrogram*

Figure 2. *Mean Plot of Lifetime CVD Risk vs. Cluster Solutions*

Figure 3. *Mean Plot of 30-Year CVD Risk vs. Cluster Solutions*