Journal of Community Health (2019) 44:500–506
https://doi.org/10.1007/s10900-018-00609-2

ORIGINAL PAPER

College Students' Perception of Current and Projected 30-Year Cardiovascular Disease Risk Using Cluster Analysis with Internal Validation

Dieu-My T. Tran[1]

Published online: 15 December 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract

Cardiovascular risk factors in young adults at a national level are less than ideal specifically for hypercholesterolemia, hypertension, and diabetes. Explore college students' perception of their 30-year cardiovascular disease (CVD) risk using cluster analysis technique with internal validation. This is a descriptive and cross-sectional study. A total of 133 college students, aged 20–36 with no known history of CVD, were recruited and used to perform cluster analysis with internal validation. The mean age of the sample was 24.85 and predominately female (59.5%). The mean score for perception of cardiovascular risk factors was 21.20 ranging from 12 to 34 points on a Likert scale. The mean score for the 30-year CVD risk assessment was 5.23 ranging from 1 to 22%. Five clusters emerged from the cluster analysis. However, two of the clusters, the highest risk with moderate perception and low risk and lowest perception, were identified as areas for potential intervention as there were discrepancies between subjects' perceived risk and their actual 30-year risk. The national data and literature has indicated a lack of awareness of CVD risk among this population which our study also concurred. Identifying the discrepancies between the perceived and projected CVD risk are useful for researchers and clinicians such as nurses to take the initiative to focus on and begin to intervene in this population to reduce potential adverse events of CVD.

Keywords Cluster analysis · Students · Cardiovascular risk · Perception

Introduction

Cardiovascular disease (CVD) is a threat to both men and women and remains the number one cause of death in the United States [3]. However, with lifestyle modifications, CVD can be controlled, treated, and prevented. Bucholz et al. [5] analyzed young adults awareness from the National Health and Nutrition Examination survey (NHANES) and found that cardiovascular risk factors in young adults at a national level are less than ideal specifically for hypercholesterolemia, hypertension, and diabetes. Early prevention is critical especially during young adulthood (18–39 years old) when young adults are gaining their independence and forming life habits that may have a substantial impact on their cardiovascular health. Not only young adults in general, but specifically young adults in college are at risk for CVD due to the college environment (i.e., eating out frequency, fast food restaurant availability, lack of physical activity, sleeping habits, stress level, alcohol consumption, and essentially food choices that are high in fats and sodium [6, 24]).

The literature has established that young adults in college are at risk of developing CVD [1, 18, 21]. It is imperative to examine risk perceptions among this population because without awareness and a congruent understanding of the risks, the chances of lifestyle modifications to optimize cardiovascular health are less likely to occur. Davidson et al. [7] examined patients that underwent a percutaneous coronary intervention and found a weak relationship between their actual and perceived risk of CVD. Therefore, examining a college population's current perception and projected 30-year CVD risk to explore the underlying patterns may be beneficial.

✉ Dieu-My T. Tran
Dieu-my.tran@unlv.edu

[1] School of Nursing, University of Nevada, Las Vegas, 4505 S. Maryland Pkwy, Box 453018, Las Vegas, NV 89154-3018, USA

Cluster Analysis Technique

Cluster analysis is a statistical technique of dividing a set of data into subsets of clusters based on the similarities and differences of variables [15]. The clustering technique is a useful method in data exploration to identify meaningful groups based on data distributions and patterns in the underlying data [11]. Cluster validation is an important component of cluster analysis to determine the true number and improve accuracy of cluster solutions. The variety of different cluster algorithms that exist produce different types of results necessary for statistical cluster validations [8]. It is often difficult to determine the exact number of clusters in a dataset because of the subjectivity and ambiguous nature of cluster analysis [4, 13–15, 23]. However, there are different approaches to determine internal validation including the seven steps of internal validation presented by Lange et al. [13].

Cluster analysis is gradually gaining popularity in research for data exploration and has been used in several disciplines [23]. However, there is a lack of studies in the literature that have used cluster analysis on cardiovascular risk factors specifically as they are an emerging population for the development of cardiovascular risk factors in college students. Studies have found that college students show a disturbingly high prevalence of cardiovascular risk factors and the majority are unaware of CVD [2, 5, 12]. In-depth evaluation of college students' perception of cardiovascular risk factors and their actual 30-year CVD risk using cluster analysis is essential to strategize effective interventions for CVD prevention. Therefore, the purpose of this study was to explore college students' perception of their 30-year CVD risk using cluster analysis technique with internal validation.

Methods

Study Sample

The present study is part of a descriptive, cross-sectional, mixed-methods study that used convenience sampling to identify and evaluate college students' cardiovascular risk factors. Subjects were recruited from a large university over 6 months. A total of 133 college students were recruited and used to perform cluster analysis with internal validation. The inclusion criteria were students at the participating university, aged 20–36 years. The exclusion criteria were college students with known history of CVD such as myocardial infarction, stroke, or coronary heart disease, currently pregnant or lactating, or weigh less than 110 pounds (to have their blood drawn). The Institutional Review Board approved the study (#953581) and informed consent was obtained from all the subjects.

Measures

An adapted version of the Health Beliefs Related to Cardiovascular Disease (HBCVD) questionnaire was used to assess the perception of cardiovascular risk factors in the studied population [19, 20]. The HBCVD questionnaire has been shown to be reliable and valid to use in the studied population [19, 20]. The adapted questionnaire consisted of two subscales pertaining to this study, perceived severity and susceptibility with ten Likert scale items ($\alpha = 0.70$), five in each subscale that focused on perceived severity ($\alpha = 0.62$) and susceptibility ($\alpha = 0.75$) [20]. Response items ranged from strongly disagree (1) to strongly agree (4). The HBCVD scores ranged from 10 to 40 points, with higher scores reflecting greater perception of cardiovascular risk factors [19, 20].

The participating university health center laboratory phlebotomists drew 10 mL of venous samples from each subject. The measures collected from all subjects included weight and height to calculate body mass index, fasting lipid panel and glucose, hemoglobin A1c, serum cotinine levels, and two resting blood pressure levels (the mean of the two resting blood pressure levels were used). The measured data from each subject were used to generate their Framingham 30-year CVD risk score [16]. Specifically, the following risk factors were included: sex (male vs. female), age (years), total cholesterol (mg/dL), high-density lipoprotein cholesterol (mg/dL), systolic blood pressure (mm Hg), treatment for high blood pressure (yes vs. no), diabetes (yes vs. no), and (h) smoking status (yes vs. no) [16].

Statistical Analysis

IBM SPSS (version 25) software was used for data entry (independently by two research assistants) and analysis. Findings were determined to be statistically significance if the p value $\leq .05$. Mean, frequency, and descriptive statistics were performed on all clusters to determine the underlying patterns of the cluster solutions. To determine if the cluster membership was significant, an ANOVA table was generated on the variables included in the cluster analysis.

Cluster Analysis with Internal Validation

The seven steps from Lange et al. [13] internal validation of the cluster solution was used to guide this study internal validation methodology. A minimum of two cluster analysis techniques were employed: hierarchical and k-means cluster analyses. Hierarchical cluster analysis was employed initially to identify the appropriate number of cluster solutions followed by the k-means cluster analysis. The variables included in the cluster analyses were transformed to a z-score to achieve standardization. Ward's method was selected for the hierarchical cluster analysis algorithm. Squared Euclidean distance was selected as the proximity measure. Two independent examiners determined the number of clusters. An examination of the scree plot generated from the distance distribution and dendrogram indicated that five clusters should be obtained based on the subgroups that clusters together during separation. All five clusters had adequate members in each group as evidence by $n > 5\%$, and the final five-cluster solution was the most appropriate outcome from the hierarchical cluster analysis.

The hierarchical cluster analysis results of five cluster solutions were used to perform the k-means cluster analysis. Based on the hierarchical and k-means cluster analyses, a line graph demonstrated their similarity of cluster solutions and the mean scores of the three variables in the analyses were included (Fig. 1; Table 1). Data demonstrated that perceived severity had the most weight (F (4) = 65.2, $p < .001$) in determining the cluster membership followed by perceived susceptibility (F (4) = 58.1, $p < .001$) and the 30-year CVD risk (F (4) = 52.9, $p < .001$). Mean profile scores from each of the clusters identified by the hierarchical and k-means cluster analyses, to establish internal validity of the cluster solution, a multi-profile multi-method correlation matrix was generated (Table 2).

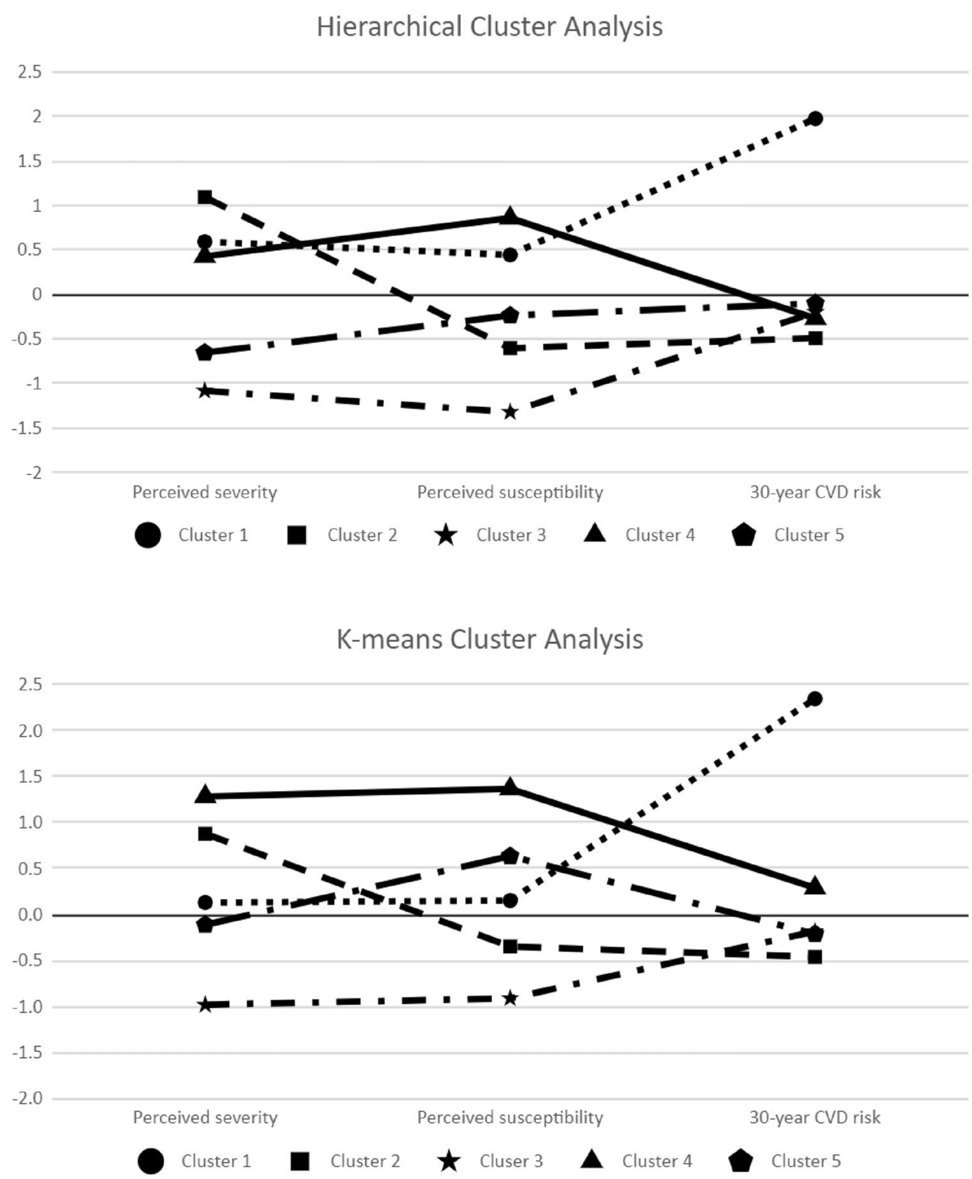

Fig. 1 Hierarchical and k-means cluster analyses five-cluster line graphs

Table 1 Mean HBCVD and 30-year CVD risk score for the hierarchical and k-means cluster solutions

Subtest	Hierarchical analysis					K-means analysis				
	Cluster H1	Cluster H2	Cluster H3	Cluster H4	Cluster H5	Cluster K1	Cluster K2	Cluster K3	Cluster K4	Cluster K5
Perceived severity	0.59	1.09	−1.08	0.43	−0.65	0.13	0.88	−0.98	1.28	−0.11
Perceived susceptibility	0.44	−0.60	−1.32	0.86	−0.24	0.15	−0.35	−0.91	1.37	0.53
30-year CVD risk	1.97	−0.49	−0.20	−0.27	−0.10	2.35	−0.46	−0.18	0.29	−0.21
N	9	17	22	47	36	6	28	43	14	40

H hierarchical cluster analysis, *K* k-means cluster analysis

Results

A total of 133 subjects met the inclusion criteria for the analysis; however, only 131 subjects were included in the final cluster analysis because there were two outliers. These two univariate outliers had the highest 30-year CVD risk at 36% and 39%, respectively, and were categorized into their own cluster. Therefore, they were excluded from the final cluster analysis. The mean age of the sample was 24.85 years (SD 4.32; range 20–36 years old) and the sample was predominately female (59.5%). The perception of cardiovascular risk factors, the HBCVD questionnaire, the mean score was 21.20 (4.03) ranging from 12 to 34 points. The 30-year CVD risk assessment mean score was 5.23 (4.09) ranging from 1 to 22%.

Five cluster solutions from the hierarchical cluster analysis retained; therefore, five clusters were used for the k-means cluster analysis. Out of the five cluster solutions from both analyses, four clusters cross-validated; however, one cluster did not cross-validate well. Despite the fact that not all of the clusters from the hierarchical and k-means cluster analyses completely cross-validated, having four out of five clusters cross-validate formed the conclusion that there was internal validity for this dataset (Fig. 1). Therefore, the k-means cluster analysis was used to interpret and determine the underlying patterns on the perception of cardiovascular risk factors and 30-year CVD risk in the studied population.

Cluster Solutions

Cluster 1: Highest Risk with Moderate Perception Cluster

Cluster 3 ($n = 6$) had the least number of subjects and they perceived themselves as moderate risk in perceived severity and susceptibility. However, this cluster had the highest 30-year CVD risk. The mean z-score for perceived severity was 0.13, perceived susceptibility was 0.15, and the 30-year CVD risk was 2.35. Subjects correctly perceived themselves at risk for CVD and their actual risk using the 30-year CVD risk assessment corresponded to having the highest 30-year CVD risk.

Cluster 2: Lowest Risk with Moderate Perception Cluster

Cluster 2 ($n = 28$) subjects had moderately high perceived severity and low perceived susceptibility with the lowest 30-year CVD risk. The mean z-score for perceived severity was 0.88, perceived susceptibility was −0.35, and the 30-year CVD risk was −0.46. Subjects perceived themselves as least likely to develop CVD, which corresponded well

Table 2 Multi-profile multi-method correlation matrix

	Hierarchical analysis					K-means analysis				
	Cluster H1	Cluster H2	Cluster H3	Cluster H4	Cluster H5	Cluster K1	Cluster K2	Cluster K3	Cluster K4	Cluster K5
Cluster H1	1.00									
Cluster H2	−0.37	1.00								
Cluster H3	0.99	−0.26	1.00							
Cluster H4	−0.96	0.08	−0.98	1.00						
Cluster H5	0.63	−0.96	0.53	−0.37	1.00					
Cluster K1	**1.00**	−0.46	0.98	−0.92	0.70	1.00				
Cluster K2	−0.49	**0.99**	−0.39	0.21	−0.99	−0.57	1.00			
Cluster K3	0.99	−0.52	**0.96**	−0.89	0.75	1.00	−0.63	1.00		
Cluster K4	−1.00	0.38	−0.99	**0.95**	−0.64	−1.00	0.51	−0.99	1.00	
Cluster K5	−0.66	−0.46	−0.74	0.85	**0.17**	−0.59	−0.33	−0.53	0.65	1.00

Bold values indicate that the hierarchical analysis procedure correlated significantly with the four corresponding profiles generated by the k-means analysis (range: r = 0.95 to r = 1.00)

H hierarchical cluster analysis, *K* k-means cluster analysis

with their actual 30-year risk assessment, as having the lowest risk among the five clusters.

Cluster 3: Low Risk and Lowest Perception

Cluster 3 had the most subjects ($n = 43$) with the lowest perception of CVD risk and surprisingly, also having a low 30-year CVD risk. The mean z-score for perceived severity was −0.98, perceived susceptibility was −0.91, and the 30-year CVD risk was −0.18. Compared to the other five clusters, these subjects had a moderate 30-year CVD risk. However, their overall 30-year CVD was low because their 30-year CVD mean risk score was less than 5%.

Cluster 4: Moderate Risk with the Highest Perception Cluster

Subjects in cluster 4 ($n = 14$) perceived themselves as the highest risk on both perceived severity and susceptibility. However, their 30-year CVD risk was only moderate compared to the other clusters. The mean z-score for perceived severity was 1.28, perceived susceptibility was 1.37, and the 30-year CVD risk was 0.29. The subjects perceived themselves as high-risk; however, their actual CVD risk was moderate.

Cluster 5: Low Risk with Low to Moderate Perception Cluster

Cluster 5 had the second highest number of subjects ($n = 40$), and they had a low perceived severity and moderate perceived susceptibility with a low 30-year CVD risk. The mean z-score for perceived severity was −0.11, perceived susceptibility was 0.63, and the 30-year CVD risk was −0.21.

Discussion

The five clusters derived from the dataset have potential implications for planning and implementing effective intervention strategies for specific subgroups of college students. Participants from Clusters 1 and 3 would be potential candidates to target for risk reduction programs because both clusters have discrepancies with their perceived risk compared to their actual 30-year risk assessment. Subjects in cluster 1 perceived themselves as having moderate risk; however, as a group, they had the highest 30-year CVD risk. Also, subjects in cluster 3 perceived their risk as lower compared to their actual 30-year CVD risk. The lack of awareness in this population of their actual CVD risk could explain this discrepancy. Bucholz et al. [5] found that 30% of young adults with diabetes, 37% with hypertension, and 43% with hypercholesterolemia were unaware of their condition [5]. This is consistent with previous findings from a nationally representative sample of young adults that 75% of those diagnosed with hypertension were not aware they had elevated blood pressure [9].

By performing internal validation on college students' perception of cardiovascular risk factors and 30-year risk assessment, five unique subgroups were characterized. While recognizing that the determinants of CVD will be multifactorial and not restricted to the perception of cardiovascular risk factors or 30-year CVD risk, it is imperative to identify key discrepancies among college students' perception and actual CVD risk. Previous studies are scarce on this topic, and the present study sought to identify the underlying patterns of college students' perception and actual CVD risk through cluster analysis, including validating the final solutions. Few studies have completed internal validation, and to the best of my knowledge, there are no studies performed on internal validation on college students' cardiovascular risk factors or their perceptions. Studies that have completed internal validation, which may not be specifically related to college students on cardiovascular risk factors, demonstrated internal validation as a useful method to increase the objectively of cluster analysis [10, 17, 22]. Specifically, Greene et al. [10] performed a two-step cluster analysis, combining sequential and hierarchical techniques, in 1689 college students and revealed a three-cluster solution that was confirmed by replication. Another large study by Wang and Biddle [22] performed hierarchical and k-means cluster analyses on 2510 young person's motivational patterns in sports and physical education and supported the validity of the clusters. The absence of internal validation studies is most likely because cluster analysis is not widely used in the health sciences including insufficient literature among this population. However, when determining the validity of cluster analysis techniques, there is sufficient literature to support the merit of this method [10, 17, 22].

Despite the fact that not all of the clusters from the hierarchical and k-means cluster analyses completely cross-validated, having four out of five clusters cross-validated formed the conclusion that there was internal validity for this dataset of college students' perception of cardiovascular risk factors and 30-year CVD risk assessment.

Limitations

This study is not without limitations. The data analyzed is descriptive and cross-sectional; therefore, generalizability is limited. One of the limitations is the small sample size with a wide age range of 20–36 years. As well, in the given dataset, determining the most accurate cluster solution and its interpretation have its limitations due to the nature of cluster analysis. However, the potential limitations of cluster analysis are minimal and the benefits are quick and easy to perform. An identified limitation of this study is the subjectivity of the results especially the appropriate number of cluster solutions to obtain. However, performing the seven steps of internal validation [13] increased the objectivity of the cluster. Another limitation is that this study is cross-sectional and would benefit from additional data collection at different time points and across various samples to allow further validation of the clusters. Lastly, additional variables could have been explored on this dataset.

Conclusions

To the best of my knowledge, no studies have examined or performed cluster analysis with internal validation of college students' perception and actual CVD risk. Despite the merit of internal validation to identify specific subgroups of college student's perception and risk of CVD with discrepancies that need to be intervened. This analysis demonstrated that cluster analysis is a valid approach to examine in the health sciences on relationships between CVD, cardiovascular risk assessment, and risk perception. The national data and literature demonstrated the lack of awareness of CVD risk in this population which our study also concurred. Researchers and clinicians such as nurses should take the initiative to focus on and begin to intervene in this population to reduce potential adverse events of CVD.

Compliance with Ethical Standards

Conflict of interest The author declared no conflicts of interest related to this manuscript.

References

1. Akhtar, P. C., Haw, S. J., Currie, D. B., Zachary, R., & Currie, C. E. (2009). Smoking restrictions in the home and secondhand smoke exposure among primary schoolchildren before and after introduction of the scottish smoke-free legislation. *Tobacco Control, 18*(5), 409–415. https://doi.org/10.1136/tc.2009.030627.
2. Arts, J., Fernandez, M. L., & Lofgren, I. E. (2014). Coronary heart disease risk factors in college students. *Advances in Nutrition, 5*(2), 177–187. https://doi.org/10.3945/an.113.005447.
3. Benjamin, E. J., Blaha, M. J., Chiuve, S. E., Cushman, M., Das, S. R., Deo, R., et al. (2017). Heart disease and stroke statistics-2017 update: A report from the american heart association. *Circulation, 135*(10), e146–e603. https://doi.org/10.1161/CIR.0000000000000485.
4. Breckenridge, J. N. (2000). Validating cluster analysis: Consistent replication and symmetry. *Multivariate Behavioral Research, 35*(2), 261–285.
5. Bucholz, E. M., Gooding, H. C., & de Ferranti, S. D. (2018). Awareness of cardiovascular risk factors in US young adults aged 18–39 years. *American Journal of Preventive Medicine, 54*(4), e67–e77.
6. 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.
7. Davidson, P. M., Salamonson, Y., Rolley, J., Everett, B., Fernandez, R., Andrew, S., et al. (2011). Perception of cardiovascular risk following a percutaneous coronary intervention: A cross sectional study. *International Journal of Nursing Studies, 48*(8), 973–978. https://doi.org/10.1016/j.ijnurstu.2011.01.012.
8. Deborah, L. J., Baskaran, R., & Kannan, A. (2010). A survey on internal validity measure for cluster validation. *International Journal of Computer Science & Engineering Survey, 1*(2), 85–102.

9. Gooding, H. C., McGinty, S., Richmond, T. K., Gillman, M. W., & Field, A. E. (2014). Hypertension awareness and control among young adults in the national longitudinal study of adolescent health. *Journal of General Internal Medicine, 29*(8), 1098–1104.
10. Greene, G. W., Schembre, S. M., White, A. A., Hoerr, S. L., Lohse, B., Shoff, S., et al. (2011). Identifying clusters of college students at elevated health risk based on eating and exercise behaviors and psychosocial determinants of body weight. *Journal of the American Dietetic Association, 111*(3), 394–400.
11. Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). On clustering validation techniques. *Journal of Intelligent Information Systems, 17*(2–3), 107–145.
12. Huang, T. T., Kempf, A. M., Strother, M. L., Li, C., Lee, R. E., Harris, K. J., et al. (2004). Overweight and components of the metabolic syndrome in college students. *Diabetes Care, 27*(12), 3000–3001.
13. Lange, T., Roth, V., Braun, M. L., & Buhmann, J. M. (2004). Stability-based validation of clustering solutions. *Neural Computation, 16*(6), 1299–1323.
14. Liu, Y., Li, Z., Xiong, H., Gao, X., & Wu, J. (2010). Understanding of internal clustering validation measures. In *Paper presented at IEEE 10th international conference on data mining (ICDM)*, pp. 911–916.
15. Liu, Y., Li, Z., Xiong, H., Gao, X., Wu, J., & Wu, S. (2013). Understanding and enhancement of internal clustering validation measures. *IEEE Transactions on Cybernetics, 43*(3), 982–994.
16. 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.
17. Quatromoni, P. A., Copenhafer, D. L., Demissie, S., D'Agostino, R. B., O'Horo, C. E., Nam, B., et al. (2002). The internal validity of a dietary pattern analysis. the framingham nutrition studies. *Journal of Epidemiology and Community Health, 56*(5), 381.
18. Racette, S. B., Deusinger, S. S., Strube, M. J., Highstein, G. R., & Deusinger, R. H. (2005). Weight changes, exercise, and dietary patterns during freshman and sophomore years of college. *Journal of American College Health, 53*(6), 245–251.
19. Tovar, E. G., Rayens, M. K., Clark, M., & Nguyen, H. (2010). Development and psychometric testing of the health beliefs related to cardiovascular disease scale: Preliminary findings. *Journal of Advanced Nursing, 66*(12), 2772–2784. https://doi.org/10.1111/j.1365-2648.2010.05443.x.
20. Tran, D. M., Zimmerman, L. M., & Kupzyk, K. A. (2016). Validation of the knowledge and perception of cardiovascular risk factors questionnaires for college students. *Journal of Nursing Measurement, 24*(2), 202–214. https://doi.org/10.1891/1061-3749.24.2.202.
21. Vella-Zarb, R. A., & Elgar, F. J. (2009). The 'freshman 5': A meta-analysis of weight gain in the freshman year of college. *Journal of American College Health, 58*(2), 161–166.
22. Wang, C. J., & Biddle, S. J. (2001). Young people's motivational profiles in physical activity: A cluster analysis. *Journal of Sport and Exercise Psychology, 23*(1), 1–22.
23. Yim, O., & Ramdeen, K. T. (2015). Hierarchical cluster analysis: Comparison of three linkage measures and application to psychological data. *The Quantitative Methods for Psychology, 11*(1), 8–21.
24. Zaki, S. M., Ajabnoor, M. A., Aziz, M. A., & Hassan, R. M. (2013). Relation of body composition to dietary habits and lifestyles among male college students with possible blood pressure affection: A cross section study. *Wulfenia Journal, 20*(4), 231–244.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.