J Korean Med Sci. 2023 May 1;38(17):e133
https://doi.org/10.3346/jkms.2023.38.e133
eISSN 1598-6357·pISSN 1011-8934

JKMS

Original Article
Humanities &
Forensic Medicine

Impact of Criterion Versus Norm-Referenced Assessment on the Quality of Life in Korean Medical Students

Ce Hwan Park ,[1*] Jihyeon Kwon ,[1*] Jong Tae Lee ,[2] and Sangzin Ahn [3]

[1]Inje University College of Medicine, Busan, Korea
[2]Department of Preventive Medicine, Inje University College of Medicine, Busan, Korea
[3]Department of Pharmacology and Pharmacogenomics Research Center, Inje University College of Medicine, Busan, Korea

OPEN ACCESS

Received: Aug 29, 2022
Accepted: Jan 11, 2023
Published online: Apr 17, 2023

Address for Correspondence:
Sangzin Ahn, MD, PhD
Department of Pharmacology, Inje University College of Medicine, 75 Bokji-ro, Busanjin-gu, Busan 47392, Korea.
Email: sangzinahn@gmail.com

*Ce Hwan Park and Ji Hyeon Kwon contributed equally to this work.

© 2023 The Korean Academy of Medical Sciences.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (https://creativecommons.org/licenses/by-nc/4.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ORCID iDs
Ce Hwan Park
https://orcid.org/0000-0002-9254-8553
Jihyeon Kwon
https://orcid.org/0000-0001-7584-2665
Jong Tae Lee
https://orcid.org/0000-0002-6132-897X
Sangzin Ahn
https://orcid.org/0000-0003-2749-0014

Funding
This research was supported by the Inje University Research Grant of 2021.

Disclosure
The authors have no potential conflicts of interest to disclose.

ABSTRACT

Background: Medical students are known to be subjected to immense stress under competitive curricula and have a high risk of depression, burnout, anxiety and sleep disorders. There is a global trend of switching from norm-referenced assessment (NRA) to criterion-referenced assessment (CRA), and these changes may have influenced the quality of life (QOL), sleep phase, sleep quality, stress, burnout, and depression of the medical students. We hypothesized that there is a significant difference of QOL between CRA and NRA and that sleep, stress, burnout, and depression are the main contributors.
Methods: By administering an online survey regarding QOL and its contributors to Korean medical students, 365 responses from 10 medical schools were recorded. To clarify the complex relationship between the multiple factors in play, we applied nonlinear machine learning algorithms and utilized causal structure learning techniques on the survey data.
Results: Students with CRA had lower scores in stress (68.16 ± 11.29, 76.03 ± 12.38, $P < 0.001$), burnout (48.09 ± 11.23, 55.93 ± 13.07, $P < 0.001$), depression (12.77 ± 9.82, 16.44 ± 11.27, $P = 0.003$) and higher scores in QOL (95.79 ± 16.20, 89.65 ± 16.28, $P < 0.001$) compared with students with NRA. Multiple linear regression, permutation importance of the random forest model and the causal structure model showed that depression, stress and burnout are the most influential factors of QOL of medical students.
Conclusion: Medical students from schools that use CRA showed higher QOL scores, as well as lower burnout, stress and depression when compared with students from schools that use NRA. These results may be used as a basis for granting justification for the transition to CRA.

Keywords: Criterion-Referenced Assessment; Medical Education; Quality of Life; Stress; Burnout; Depression

INTRODUCTION

The curriculum of medical school is inherently stressful. Medical students have a high risk of depression, burnout, anxiety, sleep disorders, and are subjected to immense stress under competitive grading systems.[1] Prior researches generally confirm that moving toward the pass/fail grading system improves students' psychological well-being (reduce stress and anxiety), increase self-efficacy, decrease competitiveness, and promote cooperative

Generated by xmlinkpress

Author Contributions

Conceptualization: Park CH, Kwon J, Ahn S. Formal analysis: Park CH, Kwon J, Ahn S. Funding acquisition: Ahn S. Investigation: Park CH, Kwon J, Lee JT. Methodology: Park CH, Kwon J, Ahn S. Software: Park CH, Kwon J, Ahn S. Validation: Park CH. Visualization: Kwon J. Writing - original draft: Park CH, Kwon J. Writing - review & editing: Lee JT, Ahn S.

learning.[2-5] Against many concerns, the change from the traditional grading system to the pass/fail system did not show reduction in the academic performance, USMLE test scores, success in residency placement, and level of attendance.[1,6] Despite the trend to introduce criterion-referenced assessment (CRA) in medical schools around the world based on these advantages and about 88% of medical schools in the United States in 2020–2021 use this evaluation method, norm-referenced assessment (NRA) is still widely used in many countries.[7] In particular, only two (5%) of forty medical schools in South Korea have adopted CRA as of 2022. Furthermore, medical students in South Korea are known to have depression, burnout, sleep problems, and high stress under a heavy-loaded medical curriculum, which may lead to low QOL.[8-11] Improved quality of life (QOL) for medical students may be essential for future career motivation and the prevention of depression, anxiety, or sleep disorder in residency.[12] However, there is insufficient comparative data available to guide decision making in this area between CRA and NRA among medical schools.

This study aimed to investigate whether CRA offers higher QOL to Korean medical students compared with NRA. Our hypothesis is that there is a significant difference of QOL between the two assessment groups alongside its 5 related factors (sleep phase, sleep quality, stress, burnout and depression). Hence, a cross-sectional survey was administered to Korean medical students. In addition to typical multiple linear regression (MLR) analyses, machine learning regression models were applied in an attempt to capture non-linear relationships between the multiple factors in play, and causal structure learning techniques were utilized to model the putative directions of effects between the factors with the given information from a single cross-sectional survey.

METHODS

Online survey

In South Korea, medical schools with a 4-year curriculum that consists of 2 years of preclerkship and 2 years of clerkship phase, and those with a 6-year curriculum that consists of 2 years of premedical, 2 years of preclerkship and 2 years of clerkship phase coexist. The preclerkship phase curriculum is based on a typical lecture-and-test style education, while the clerkship phase consists of various methods of learning and assessment, depending on the clerkship hospital, department, subject or school. Hence, the study target was limited to students in their preclerkship phase. Students in their 2 years of preclerkship phase from 10 universities were invited for the online survey, including 2 schools which use CRA: Inje University, College of Medicine; Yonsei University, College of Medicine and 8 schools which use NRA: Hallym University, College of Medicine; Hanyang University, College of Medicine; Keimyung University, College of Medicine; Korea University, College of Medicine; Kyungpook National University, School of Medicine; Pusan National University, School of Medicine; Seoul National University, College of Medicine; The Catholic University of Korea, College of Medicine. The survey included 6 validated questionnaires that yield a numerical score regarding sleep (reduced Morningness–Eveningness Questionnaire [rMEQ], 5 to 23 points; Pittsburgh Sleep Quality Index [PSQI], 0 to 21 points), stress (Medical Student Stress Inventory in Korea [MSSIK], 9 to 116 points), burnout (Korean Maslach Burnout Inventory Student Survey [K-MBI-SS], 14 to 98 points), depression (Center for Epidemiologic Studies Depression Scale [CESD], 0 to 60 points) and QOL (QOL assessment developed by the WHOQOL group, 26 to 130 points).[13-18] MSSIK is a questionnaire devised to measure stress while considering the sociocultural environment in Korean medical schools. Since our survey targeted students in their preclerkship phase, we excluded 11 questions related to the stress of clinical years from the MSSIK questionnaire.[15] K-MBI-SS is a modified version of the MBI-SS which has been validated with Korean students in previous studies.[16]

The survey was administered using Google Forms platform starting on October 22, 2021 and closing on November 16, 2021. The web address to the survey was distributed to student representatives of each school and was conveyed to students using closed membership social media groups or messenger services.

Average comparison and linear regression

Data validation was processed by Shapiro-Wilk test, Durbin-Watson test, Non-constant Variance Score test, Bonferroni Outlier test and testing for multicollinearity with variance inflation factor. The respondents were divided into the students with CRA (CRA group) and the students with NRA (NRA group). The χ^2 test was performed to determine potential differences in gender or grade (preclerkship year 1 or 2) across groups. Student's *t*-test was implemented to check for differences of the scores of five factors and QOL between gender groups. The mean and standard deviation of each score of two groups were compared through the aspect of 6 domains including sleep phase, sleep quality, stress, burnout, depression and QOL. The QOL scores were further divided into four domains: physical, psychological, social and environmental.[19,20] MLR was applied to analyze the combined relationship between QOL (independent variable) and the 5 factors (dependent variables).

Machine learning regression model training and validation

An array of 7 base models were developed with nonlinear machine learning regression algorithms including k-nearest neighbors, decision tree, epsilon-support vector regression, random forest (RF), AdaBoost (ADA), gradient boosting machine and multilayer perceptron (ANN). The base models were trained with the scikit-learn package (version 1.0.2) using default settings and 5-fold cross validation, except for setting ANN to use a single hidden layer of 5 nodes due to the small input data dimension. The top 3 models with the highest R squared value among the 7 base models were selected and further trained and validated using nested 5-fold cross validation and hyperparameter tuning with random search. Eventually, RF showed the highest R squared value after hyperparameter tuning, and thus was opted as the final nonlinear machine learning model. The MLR model was also trained and validated using 5-fold cross validation to predict QOL scores with scores of the 5 factors. The 5-fold cross validation strategy was incorporated to measure the accuracy of predictions made with observations not included in the training process, while coping with the bias of random train-test splitting. Student's *t*-test was applied to the 5-fold cross validation R^2 results to check for difference in performance of the two models.

Feature contribution and importance scores

Individual conditional expectations (ICEs) are predictions (y axis) of each data when a feature (x axis) changes which can be expressed as a line per one instance (one student). The value for each line is calculated by keeping all other variables the same as the original data, but changing the value of one specific variable across a range and making predictions with the model. This visualizes the dependence of the predictions of a model on one specific independent variable. Partial dependence plot (PDP) is the average of all ICEs and represents the global effect of an independent variable on the predictions of a model.[21] PDP and ICEs were yielded by training a model on the entire dataset and applying the PartialDependenceDisplay function from the scikit-learn package.

In order to quantify the contribution of each feature in a trained model, permutation importance scores were calculated using the permutation_importance function from the scikit-learn package. Briefly, after randomly shuffling a single feature, the model makes new predictions with the permuted dataset and determines the decrease in R^2 value when compared with the predictions made with the original data. A larger reduction of performance indicates a higher importance of the feature. The random permutations were performed with 30 repeats.

Causal structure learning

With the aim to analyze the direction of effect between each questionnaire scores, the Non-combinatorial Optimization via Trace Exponential and Augmented lagRangian for Structure learning (NOTEARS) algorithm was applied to the standard z-scores of groups (coded as 0 for norm-based, 1 for criterion-referenced) and the 6 questionnaire scores, using the causalnex package (version 0.11.0). NOTEARS is a machine learning algorithm that learns a directed acyclic graph (DAG) which describes the conditional dependencies between multiple variables for the establishment of Bayesian Belief Network.[22] Each node of the DAG represents a random variable, and the weight of each edge that directs from a parent node to a child node means that x standard deviation change in the parent variable causes x × weight standard deviation change in the child variable. The weight value represents the size of effect, not the robustness of directionality. As the NOTEARS algorithm allows knowledge-based user inputs to the graph, the authors applied constraints that Group cannot be a child node, and that QOL cannot be a parent node. These constraints are based on simple logical assumptions, such as questionnaire results cannot have a causal effect on the student's method of academic assessment, and the QOL of a respondent is influenced by other factors but does not have causal effect on other aspects of a student. The absolute edge weight threshold value was set to 0.08.

Analytic tools and software

Data validation and preprocessing was performed with RStudio version 1.4 using R 4.1.2. Other analyses and data visualizations were performed on Google Colab platform using Python 3.6.9. Machine learning models were trained and analyzed using the scikit-learn package (version 1.0.2), and the MLR was performed using the statsmodels package (version 0.3.1). Spearman correlation and heatmap was produced using scipy (version 1.7.3) and seaborn (version 0.11.2) libraries.

Ethics statement

The study protocol was approved by the Institutional Review Board (IRB) of Inje University Busan Paik Hospital (IRB number: 2021-09-034). Informed consent was obtained electronically from all participants at the beginning of the online survey.

RESULTS

Baseline characteristics

A total of 365 medical students agreed and participated in our survey. In the CRA group, 111 students participated in the survey and in the NRA group, 254 students participated. Eight responses were excluded because a respondent from the CRA group answered the NRA group questionnaire or the response was a duplicate. The response rate of CRA group students was 23.9% (111 responses from 465 students) and that of NRA group students was 16.3% (254 responses from 1,563 students). The result of the χ^2 test in grade (1 and 2 year of preclerkship) was not significant ($P = 0.657$) while that in gender was significant ($P = 0.016$) (**Table 1**). There were no significant differences between male and female in rMEQ (10.90 ± 3.10, 10.93 ± 3.43, $P = 0.863$), PSQI (6.67 ± 2.66, 6.57 ± 2.60, $P = 0.845$), MSSIK (73.20 ± 12.58, 74.14 ± 12.59, $P = 0.481$), K-MBI-SS (54.04 ± 13.25, 52.97 ± 12.80, $P = 0.252$), CESD (15.01 ± 11.16, 15.69 ± 10.76, $P = 0.364$) and QOL (mean 91.88 ± 16.30, 91.10 ± 16.72, $P = 0.652$), hence further analysis was performed without weight adjustment. The correlation matrix between assessment group, gender, grade, 5 domains and QOL is provided in **Supplementary Fig. 1**.

Comparison of CRA and NRA groups on QOL and the 5 related domains (sleep phase, sleep quality, stress, burnout and depression)

When comparing scores of QOL and its related factors among students, the CRA group showed higher QOL scores compared with the NRA group (95.79 ± 16.20, 89.65 ± 16.28, $P < 0.001$). MSSIK (stress scale, 68.16 ± 11.29, 76.03 ± 12.38, $P < 0.001$), K-MBISS (burnout scale, 48.09 ± 11.23, 55.93 ± 13.07, $P < 0.001$), and CESD (depression scale, 12.77 ± 9.82, 16.44 ± 11.27, $P = 0.003$) scores in CRA group were significantly lower than those in NRA group. There was no significant difference in rMEQ (sleep phase scale, 10.72 ± 3.62, 11.00 ± 3.08, $P = 0.469$) and PSQI (sleep quality scale, 6.41 ± 2.70, 6.72 ± 2.60, $P = 0.287$) scores between the two groups (**Table 2**).

Table 1. Comparison of gender, grade, depression, four QOL domains between the students with criterion-referenced assessment and the students with norm-referenced assessment

Domains	All students (n = 365)	Students with criterion-referenced assessment (n = 111)	Students with norm-referenced assessment (n = 254)	P value[a]
Gender				0.016
Male	196 (53.70)	49 (44.14)	147 (57.87)	
Female	169 (46.30)	62 (55.86)	107 (42.13)	
Grade				0.657
Year 1	151 (41.37)	44 (39.64)	107 (42.13)	
Year 2	214 (58.63)	67 (60.36)	147 (57.87)	
Depression				0.024
CESD ≥ 16	154 (42.20)	37 (33.33)	117 (46.06)	
CESD < 16	211 (57.80)	74 (66.67)	137 (53.94)	
QOL				
Physical	14.72 ± 2.76	15.53 ± 2.70	14.37 ± 2.72	< 0.001
Psychological	13.64 ± 3.04	14.20 ± 2.76	13.39 ± 3.13	0.029
Social	13.63 ± 3.42	14.34 ± 3.20	13.32 ± 3.48	0.015
Environmental	14.52 ± 2.85	15.01 ± 2.79	14.31 ± 2.86	0.047

Values are presented as number (%) or mean ± standard deviation.
QOL = quality of life, CESD = Center for Epidemiologic Studies Depression Scale.
[a]Chi-squared test was used for categorical values and *t*-test was used for continuous values.

Table 2. Average comparison between the score of QOL and five related factors in students with criterion-referenced assessment and students with norm-referenced assessment

Scales	Mean ± standard deviation			P value
	All students (n = 365)	Students with criterion-referenced assessment (n = 111)	Students with norm-referenced assessment (n = 254)	
rMEQ (sleep phase)	10.92 ± 3.25	10.72 ± 3.62	11.00 ± 3.08	0.469
PSQI (sleep quality)	6.63 ± 2.63	6.41 ± 2.70	6.72 ± 2.60	0.287
MSSIK (stress)	73.64 ± 12.57	68.16 ± 11.29	76.03 ± 12.38	< 0.001
K-MBI-SS (burnout)	53.54 ± 13.03	48.09 ± 11.23	55.93 ± 13.07	< 0.001
CESD (depression)	15.32 ± 10.97	12.77 ± 9.82	16.44 ± 11.27	0.003
WHOQOL-BREF (QOL)	91.52 ± 16.48	95.79 ± 16.20	89.65 ± 16.28	< 0.001

QOL = quality of life, rMEQ = reduced Morningness–Eveningness Questionnaire, PSQI = Pittsburgh Sleep Quality Index, MSSIK = Medical Student Stress Inventory in Korea, K-MBI-SS = Korean Maslach Burnout Inventory Student Survey, CESD = Center for Epidemiologic Studies Depression Scale, WHOQOL-BREF = quality of life assessment developed by the WHOQOL group.

As a result of comparing the QOL in four domains, scores of CRA group were significantly higher than those of NRA group in all domains including physical (15.53 ± 2.70, 14.37 ± 2.72), psychological (14.20 ± 2.76, 13.39 ± 3.13), social (14.34 ± 3.20, 13.32 ± 3.48), environmental (15.01 ± 2.79, 14.31 ± 2.86) QOL. Students with a CESD score of 16 or higher were more common in the NRA group among the two groups (42.20%, 33.33%, $P = 0.024$).

Correlations between 5 factors and QOL

Results of a MLR between the 5 factors and QOL are shown in **Table 3**. When analyzed with all students, MSSIK and CESD showed a significant negative correlation (MSSIK: $\beta = -0.2079$, $P < 0.001$; CESD: $\beta = -0.5170$, $P < 0.001$) with the QOL score. K-MBI-SS also showed a negative correlation, but did not reach significance ($P = 0.061$). Sleep-related variables showed an insignificant correlation with the QOL score. In the CRA group, K-MBI-SS and CESD showed a significant negative correlation (K-MBI-SS: $\beta = -0.2429$, $P = 0.028$; CESD: $\beta = -0.4074$, $P < 0.001$) with the QOL score. In NRA group, MSSIK and CESD showed a significant negative correlation (MSSIK: $\beta = -0.2488$, $P < 0.001$; CESD: $\beta = -0.5512$, $P < 0.001$) with the QOL score, similar with the correlations observed when analyzed with all students (**Table 3**).

Comparison of the linear regression and machine learning model for predicting factors related to QOL

The three nonlinear models ANN, ADA and RF showed the highest R^2 values among the 7 base models. After performing hyperparameter tuning with the 3 models, RF was selected as the final nonlinear model since it had the greatest R^2 value (RF: 0.41 ± 0.10, ANN: 0.38 ± 0.09, ADA: 0.39 ± 0.09). The performance of the MLR model and RF model measured by R^2 with 5-fold cross validation was not statistically different (0.44 ± 0.11, 0.41 ± 0.10, $P = 0.621$), indicating that MLR and RF had similar performances in predicting QOL scores.

Depression (CESD, 0.541 ± 0.045 in logistic regression [LR], 0.583 ± 0.036 in RF), stress (MSSIK, 0.105 ± 0.017 in LR, 0.154 ± 0.012 in RF), and burnout (K-MBI-SS, 0.020 ± 0.006 in LR, 0.129 ± 0.009 in RF) displayed the highest permutation feature importance among the 5 factors. The permutation importance in depression was notably higher compared

Table 3. Multiple linear regression in predicting quality of life among medical students with five factors

Groups	Variable	Unstandardized coefficients		Standardized coefficient	*t*	*P* value
		SE	B	Beta		
All students	rMEQ	0.178	0.1971	0.0389	1.110	0.268
	PSQI	0.244	−0.3418	−0.0545	−1.402	0.162
	MSSIK	0.065	−0.2725	−0.2079	−4.206	< 0.001
	K-MBI-SS	0.067	−0.1270	−0.1005	−1.882	0.061
	CESD	0.069	−0.7767	−0.5170	−11.217	< 0.001
Students with criterion-referenced assessment	rMEQ	0.323	0.2116	0.0472	0.656	0.513
	PSQI	0.519	−0.3910	−0.0651	−0.753	0.453
	MSSIK	0.138	−0.1194	−0.0833	−0.866	0.389
	K-MBI-SS	0.157	−0.3502	−0.2429	−2.228	0.028
	CESD	0.166	−0.6719	−0.4074	−4.039	< 0.001
Students with norm-referenced assessment	rMEQ	0.323	0.2116	0.0463	0.656	0.263
	PSQI	0.519	−0.3910	−0.0562	−0.753	0.204
	MSSIK	0.138	−0.1194	−0.2488	−0.866	< 0.001
	K-MBI-SS	0.157	−0.3502	−0.0449	−2.228	0.458
	CESD	0.166	−0.6719	−0.5512	−4.039	< 0.001

SE = standard error, rMEQ = reduced Morningness–Eveningness Questionnaire, PSQI = Pittsburgh Sleep Quality Index, MSSIK = Medical Student Stress Inventory in Korea, K-MBI-SS = Korean Maslach Burnout Inventory Student Survey, CESD = Center for Epidemiologic Studies Depression Scale.

with the other factors in both models, suggesting depression to be the most important factor contributing to QOL scores (**Fig. 1**). In the PDP and ICEs of the RF model a nonlinear relationship was observed in CESD, while the remaining four factors showed similar patterns with the linear regression model. From CESD score 0 to 20, QOL decreased drastically as CESD scores increased. But from more than 20, QOL did not react as sensitively to the rise in CESD scores (**Fig. 1B**). In the case of the linear regression model the QOL scores steadily decreased across all value ranges of the CESD score (**Fig. 1A**), which is a typical result of a linear model.

The causal structure model of QOL

In the DAG depicting the causal structure between each feature, a change in the group node, which means transition from NRA to CRA was indicated to cause an evening type sleep phase (weight = −0.083), reduction of burnout (weight = −0.081) and stress (weight = −0.288) scores. Increases in stress scores were shown to raise the burnout (weight = 0.681) and depression scores (weight = 0.199), and also have a direct influence on reducing QOL scores (weight = −0.207). Higher burnout scores were illustrated to worsen sleep quality (weight = 0.096), depression (weight = 0.480) and directly lower QOL scores (weight =

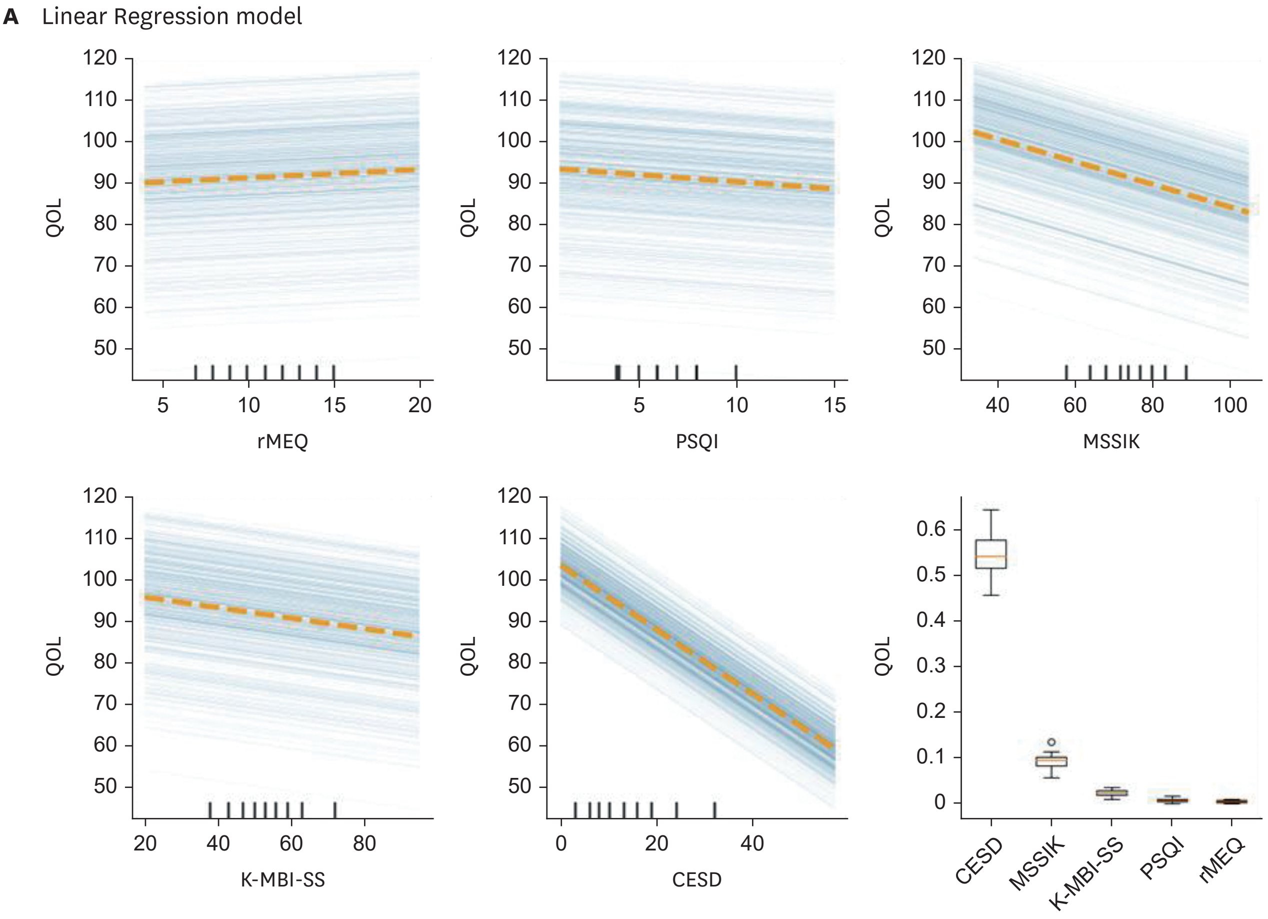

Fig. 1. Partial dependence plots (dotted line) and individual conditional expectations (solid lines) in (**A**) linear regression model and (**B**) random forest model. Boxplots present the permutation importance of each independent variable.
QOL = quality of life, rMEQ = reduced Morningness–Eveningness Questionnaire, PSQI = Pittsburgh Sleep Quality Index, MSSIK = Medical Student Stress Inventory in Korea, K-MBI-SS = Korean Maslach Burnout Inventory Student Survey, CESD = Center for Epidemiologic Studies Depression Scale.

(continued to the next page)

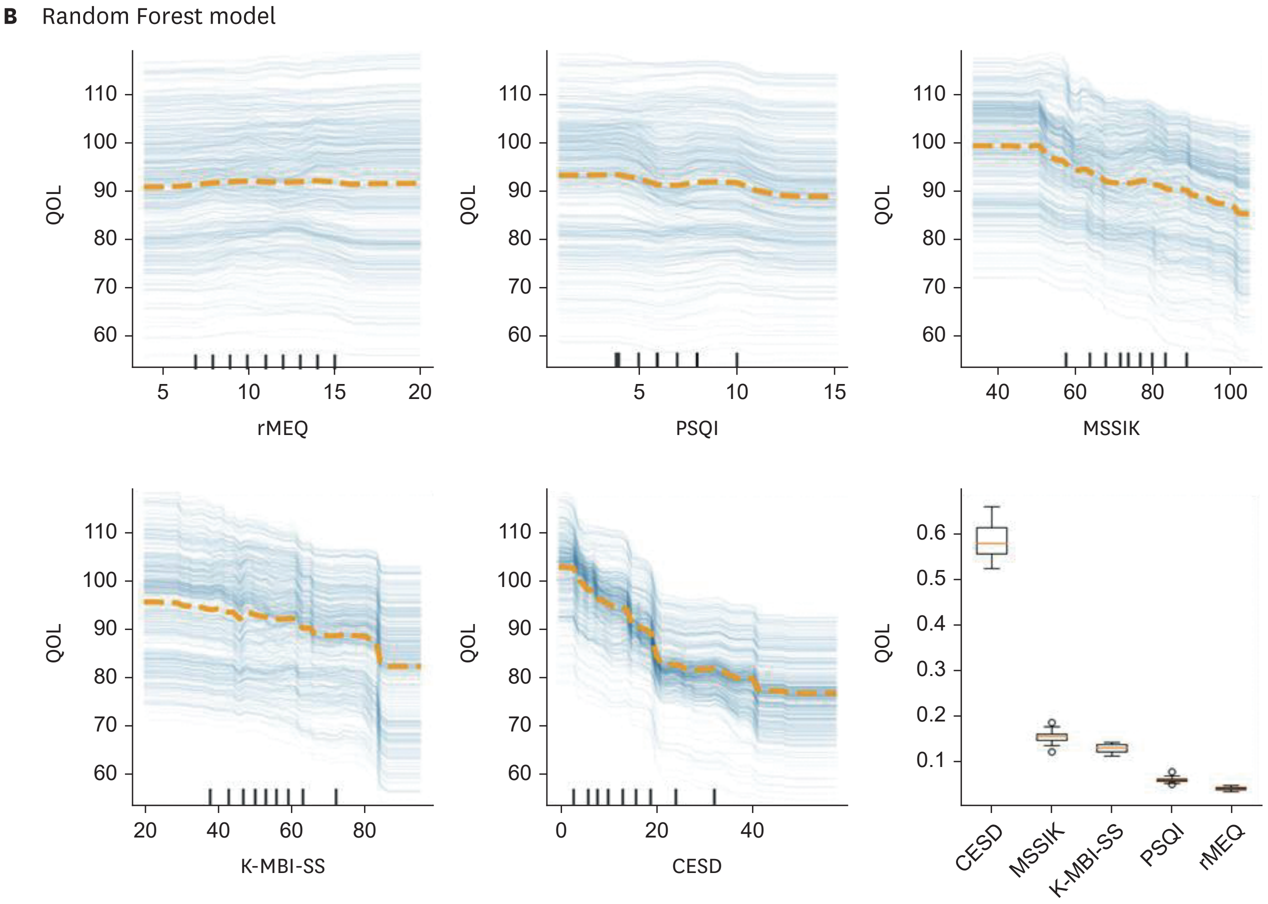

Fig. 1. (Continued) Partial dependence plots (dotted line) and individual conditional expectations (solid lines) in (**A**) linear regression model and (**B**) random forest model. Boxplots present the permutation importance of each independent variable.
QOL = quality of life, rMEQ = reduced Morningness–Eveningness Questionnaire, PSQI = Pittsburgh Sleep Quality Index, MSSIK = Medical Student Stress Inventory in Korea, K-MBI-SS = Korean Maslach Burnout Inventory Student Survey, CESD = Center for Epidemiologic Studies Depression Scale.

−0.100). Furthermore, escalation in depression scores was shown to deteriorate sleep quality (weight = 0.337) as well as QOL scores (weight = −0.517). To recapitulate, stress, burnout and depression had direct effects on QOL, and the indirect effect of assessment type was propagated mainly through stress, burnout, depression and eventually to QOL (**Fig. 2**). The factors that have the highest Spearman correlation coefficients with group were stress (r = −0.295, $P < 0.001$) and burnout (r = −0.290, $P < 0.001$), which had direct connections in the model (**Supplementary Fig. 1**). Adding gender and grade features to the model showed minimal change of the DAG structure, suggesting the robustness of the proposed model (**Supplementary Fig. 2**).

DISCUSSION

Our study suggests that CRA could improve the QOL of medical students. Interestingly, the CRA group showed higher scores of QOL and lower scores of stress, burnout and depression compared with the NRA group. In this study, MLR and machine learning were utilized for a comprehensive evaluation of multiple domains that are related to QOL. MLR and the RF model results showed that stress, burnout and depression scores had a negative correlation

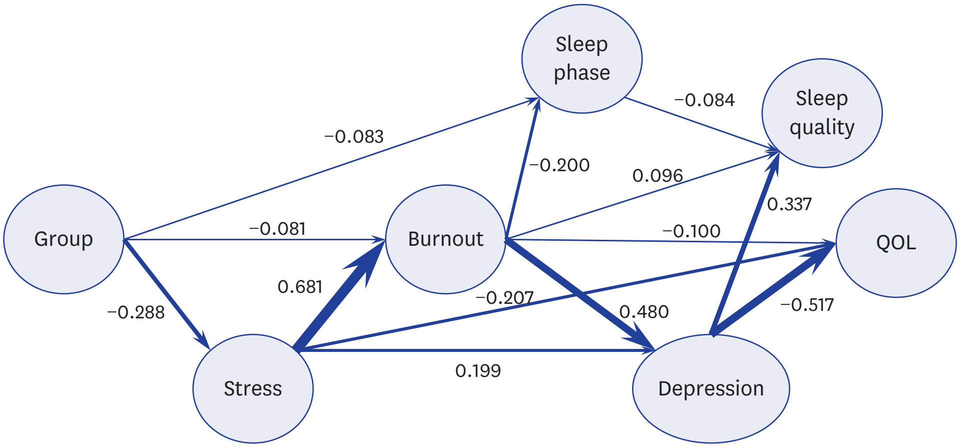

Fig. 2. Directed acyclic graph depicting causal structure network between group, five factors and QOL. Group is a variable that divides participants into two groups by the grading system (criterion-referenced assessment, norm-referenced assessment). Numbers on each edge indicate the weight value of the causal relationship.
QOL = quality of life.

with QOL scores, and the causal structure learned from the data also implied that stress, burnout and depression were the factors that had direct causal relationship with QOL.

To the authors best knowledge, this is the first multi-centered study to compare the QOLs of Korean medical students between schools that adapt CRA and NRA. The results that NRA students have higher stress, burnout, depression as well as lower QOL were similar to the previous studies that revealed that students in schools using pass/fail grading had less stress, burnout, depersonalization and anxiety compared with students in schools using 3 or more grades.[1,2] In terms of depression, there was a noticeable difference in CESD scores between two groups. The average CESD score of the NRA group was 16.44 ± 11.27 and the portion of students with a score of 16 or higher was 46.1%, which is substantially higher than the global prevalence of depression in medical students from a meta-analysis by Puthran et al. (28.0%)[23] and a previous result that surveyed Korean medical students in their preclerkship school years by Jeong et al. (14.1 ± 8.6).[24] Albeit the coronavirus disease 2019 (COVID-19) pandemic situation may invalidate direct comparisons with preceding reports, these results suggest that a substantial proportion of medical students in this group may need assessment and interventions regarding depression.

In the causal structure model, stress, burnout and depression were the factors that directly affected QOL. These results are similar with previous studies showing that high levels of stress, burnout, and depression are associated with poor QOL.[25-29] Another noteworthy point in the causal structure model is the sequence of effects, which was in the order of stress, burnout, depression and QOL. Stress was proposed as a precursor of worse burnout in medical students in previous studies, which is in line with these results.[30,31] Burnout had the smallest direct effect on QOL, but had a large effect on depression, which may be the mechanistic explanation of how burnout plays a large role in determining QOL. These results are also in consonance with the multiple reports that demonstrate that depression is the most influential estimator of poor QOL in Koreans.[28,32,33]

Sleep phase and sleep quality were not significantly different between two groups, and the multiple regression results concluded the effects to be insignificant. The causal structure learning results also did not show a direct causal relationship between the 2 sleep related variables and QOL. Both groups showed on average an evening type sleep phase and low sleep quality. This is similar to the results of a previous study and is also related to the behavior of Korean medical students which tend to study late at night.[8,34]

In an earlier study, the subjective amplitude of circadian rhythm was a more influential factor to QOL, compared with the sleep phase.[35] This may suggest that sensitivity to natural environmental synchronizers is a separate individual trait shaping the medical students' tendency to energy and mood changes, which may be related to worse QOL. However, rMEQ was developed to measure the sleep phase from circadian rhythm, not the subjective amplitude of circadian rhythm. Therefore, the sleep phase scores gauged by rMEQ may not have been an influential factor to QOL in our study.

In the case of PSQI, a possible explanation is that medical students may sleep less or adjust their sleep schedule in order to study more and be rewarded with better test results, or attend extracurricular activities which can in turn increase QOL.[36] Another explanation is that there is an overlap of questions between PSQI and CESD questionnaires, thus PSQI may partially reflect the negative cognitive viewpoints and pessimistic thinking in lieu of enclosing objectively observed aspects of sleep.[37]

To capture nonlinear correlations between QOL and the five factors, a RF model was compared with a classical MLR model. Although the RF model did not have significant difference in predictive performance and permutation importance scores showed a similar pattern with the MLR model, a nonlinear relationship was observed between depression and QOL. The steep decline in QOL in the range of low CESD scores (**Fig. 1B**) suggests that active interventions to lower medical students' depression, even in the case where the student does not satisfy conventional cutoffs, may be effective in improving the QOL.

A causal structure learning method was also employed to help better explain the correlations observed in a cross-sectional study in the scope of direction of effects. The causal model offers a method to represent the network of variables in an intuitive way and also provides means to simulate an intervention in the network using the "do" operator.[38] The learned graph model does not verify the causal effects but provides a putative model that best explains the provided cross-sectional data. In the DAG, the effect of difference in assessment methods on QOL was explained as an indirect effect transmitted through stress, burnout and depression. These 3 factors were significantly different between the groups and also had the largest beta coefficients in the MLR analysis. In contrast, sleep phase and sleep quality were mainly affected by burnout and depression but did not influence other factors. These results provide a comprehensive explanation of the effects of academic assessment methods and the 5 factors on QOL in a logically reasonable way.

This study had several limitations. Firstly, among the 2,018 medical students that were invited, only 365 of the students responded. The low response rate (18.0%) indicates that a selection bias may be in play, where students with a certain trait may have tended to respond more or less to the survey. However, this response rate was comparable to similar studies conducted on medical students (Morgan et al.[39]: 12.5%, Cecil et al.[40]: 13.0%), and to the authors best knowledge, this was the largest comparative study inspecting the QOL and related factors in Korean medical students. Next, the performance of the MLR and nonlinear machine learning model were not satisfactorily achieved ($R^2 = 0.44, 0.41$). This may be due to failure to include other confounding factors that affect QOL. Lastly, since this study was conducted as a cross-sectional survey during the second year of the COVID-19 pandemic, the results may have been heavily affected by the changes in methods of learning and general daily lives.[41] Nonetheless, under the circumstances of continuous emergence of new severe acute respiratory syndrome coronavirus 2 variants, the transformations in learning and daily lives may not be temporary and data captured during the pandemic may have its significance.

In conclusion, CRA may improve the QOL of medical students compared with NRA through reducing depression, stress and burnout. Among the 5 factors (sleep phase, sleep quality, stress, burnout, and depression) related to QOL, depression was the most significantly associated factor with lower QOL in medical students. Schools need to be more proactive in managing students' depression, stress, and burnout, and also consider shifting from NRA to CRA for the improvement of QOL of medical students.

ACKNOWLEDGMENTS

The authors would like to thank Eunbae Yang from Department of Medical Education, Yonsei University College of Medicine for the aid in performing the survey and the constructive comments. The authors would also like to thank all medical students who have participated in this study for taking the time and effort to complete the survey.

SUPPLEMENTARY MATERIALS

Supplementary Fig. 1

Correlation heatmap between assessment group, gender, school year, 5 factors and quality of life. Numbers represent Spearman correlation coefficients. Group is coded as: 0 = norm-referenced assessment, 1 = criterion-referenced assessment; and gender is coded as: 0 = male, 1 = female. Hence, positive correlation values indicate higher values in criterion-referenced assessment group students and female students, respectively.

Click here to view

Supplementary Fig. 2

Directed acyclic graph when modelling is performed including gender and school year. In comparison to the results depicted in **Fig. 2**, additional analyses were performed while including gender and school year features. Weight threshold values were set to (**A**) 0.1 and (**B**) 0.08. Smaller threshold value settings make the graph to include edges with smaller effects. In case of threshold set to 0.1, Grade and gender were both modelled as isolated nodes. When set to 0.08, higher grade was illustrated to reduce depression. Both results were similar with the original model (**Fig. 2**), suggesting the robustness of the model with the main effects of assessment type propagating through stress, burnout, depression and eventually to QOL.

Click here to view

REFERENCES

1. Bloodgood RA, Short JG, Jackson JM, Martindale JR. A change to pass/fail grading in the first two years at one medical school results in improved psychological well-being. *Acad Med* 2009;84(5):655-62.
   PUBMED | CROSSREF
2. Reed DA, Shanafelt TD, Satele DW, Power DV, Eacker A, Harper W, et al. Relationship of pass/fail grading and curriculum structure with well-being among preclinical medical students: a multi-institutional study. *Acad Med* 2011;86(11):1367-73.
   PUBMED | CROSSREF
3. Kim IS, Yang EB, Jeon WT. Analysis of internal factors affecting learning of medical student in criterion-referenced assessment system. *Korean J Educ Res* 2015;53(4):283-303.
4. Swan Sein A, Rashid H, Meka J, Amiel J, Pluta W. Twelve tips for embedding assessment for and as learning practices in a programmatic assessment system. *Med Teach* 2021;43(3):300-6.
   PUBMED | CROSSREF
5. Spring L, Robillard D, Gehlbach L, Simas TA. Impact of pass/fail grading on medical students' well-being and academic outcomes. *Med Educ* 2011;45(9):867-77.
   PUBMED | CROSSREF
6. Rohe DE, Barrier PA, Clark MM, Cook DA, Vickers KS, Decker PA. The benefits of pass-fail grading on stress, mood, and group cohesion in medical students. *Mayo Clin Proc* 2006;81(11):1443-8.
   PUBMED | CROSSREF
7. Association of American Medical Colleges. Grading systems use by US medical schools. https://www.aamc.org/data-reports/curriculum-reports/interactive-data/grading-systems-use-us-medical-schools. Updated 2022. Accessed August 23, 2022.
8. Chang HK, Lee SJ, Park CS, Kim BJ, Lee CS, Cha B, et al. Association between quality of life and eveningness as well as sleep quality among medical students. *Sleep Med Psychophysiol* 2015;22(2):64-9.
   CROSSREF
9. Park BS, Park KH. Quality of life and its related factors between medical students and non medical students. *Asia Pac J Multimed Serv Converg Art Humanit Sociol* 2017;7(9):577-88.
   CROSSREF
10. Yune S, Im S, Lee SY, Baek SY, Kam BS. Relationships among test anxiety, academic burnout, resilience, and academic achievement of medical school students. *J Educ Innov Res* 2018;28(4):173-88.
    CROSSREF
11. Kim B, Roh H. Depressive symptoms in medical students: prevalence and related factors. *Korean J Med Educ* 2014;26(1):53-8.
    PUBMED | CROSSREF
12. Dyrbye LN, West CP, Satele D, Boone S, Tan L, Sloan J, et al. Burnout among U.S. medical students, residents, and early career physicians relative to the general U.S. population. *Acad Med* 2014;89(3):443-51.
    PUBMED | CROSSREF
13. Loureiro F, Garcia-Marques T. Morning or evening person? Which type are you? Self-assessment of chronotype. *Pers Individ Dif* 2015;86:168-71.
    CROSSREF
14. Sohn SI, Kim DH, Lee MY, Cho YW. The reliability and validity of the Korean version of the Pittsburgh Sleep Quality Index. *Sleep Breath* 2012;16(3):803-12.
    PUBMED | CROSSREF
15. Kim MJ, Park KH, Yoo HH, Park IB, Yim J. Development and validation of the medical student stress scale in Korea. *Korean J Med Educ* 2014;26(3):197-208.
    PUBMED | CROSSREF
16. Lee S, Lee D. Validation of the MBI-SS scales-based on medical school students. *Asian J Educ* 2013;14(2):165-87.
    CROSSREF
17. Chon KK, Choi SC, Yang BC. Integrated adaptation of CES-D in Korea. *Korean J Health Psychol* 2001;6(1):59-76.
18. Min SK, Lee CI, Kim KI, Suh SY, Kim DK. Development of Korean version of WHO quality of life scale abbreviated version (WHOQOL-BREF). *J Korean Neuropsychiatr Assoc* 2000;39(3):571-9.
19. Krägeloh CU, Henning MA, Hawken SJ, Zhao Y, Shepherd D, Billington R. Validation of the WHOQOL-BREF quality of life questionnaire for use with medical students. *Educ Health (Abingdon)* 2011;24(2):545.
    PUBMED
20. World Health Organization. *WHOQOL-BREF: Introduction, Administration, Scoring and Generic Version of the Assessment: Field Trial Version, December 1996*. Geneva, Switzerland: World Health Organization; 1996.

21. Goldstein A, Kapelner A, Bleich J, Pitkin E. Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. *J Comput Graph Stat* 2015;24(1):44-65.
CROSSREF

22. Zheng X, Aragam B, Ravikumar P, Xing EP. Dags with no tears: continuous optimization for structure learning. *Adv Neural Inf Process Syst*. Forthcoming 2023. DOI: 10.48550/arXiv.1803.01422.
CROSSREF

23. Puthran R, Zhang MW, Tam WW, Ho RC. Prevalence of depression amongst medical students: a meta-analysis. *Med Educ* 2016;50(4):456-68.
PUBMED | CROSSREF

24. Jeong Y, Kim JY, Ryu JS, Lee KE, Ha EH, Park H. The associations between social support, health-related behaviors, socioeconomic status and depression in medical students. *Epidemiol Health* 2010;32:e2010009.
PUBMED | CROSSREF

25. Gupchup GV, Borrego ME, Konduri N. The impact of student life stress on health related quality of life among doctor of pharmacy students. *Coll Stud J* 2004;38(2):292-302.

26. Song WI, Hwang JW, Oh ES. Relations among leisure limitation, coping stress, and quality of life in college students. *J Korea Soc Wellness* 2016;11(3):159-67.
CROSSREF

27. Pagnin D, de Queiroz V. Influence of burnout and sleep difficulties on the quality of life among medical students. *Springerplus* 2015;4(1):676.
PUBMED | CROSSREF

28. Kim JW, Han DH, Lee YS, Min KJ, Park JY, Lee K. The effect of depression, anxiety, self-esteem, temperament, and character on life satisfaction in college students. *J Korean Neuropsychiatr Assoc* 2013;52(3):150-6.
CROSSREF

29. Arslan G, Ayranci U, Unsal A, Arslantas D. Prevalence of depression, its correlates among students, and its effect on health-related quality of life in a Turkish university. *Ups J Med Sci* 2009;114(3):170-7.
PUBMED | CROSSREF

30. Miranda-Ackerman RC, Barbosa-Camacho FJ, Sander-Möller MJ, Buenrostro-Jiménez AD, Mares-País R, Cortes-Flores AO, et al. Burnout syndrome prevalence during internship in public and private hospitals: a survey study in Mexico. *Med Educ Online* 2019;24(1):1593785.
PUBMED | CROSSREF

31. Zhou W, Pu J, Zhong X, Yang W, Teng T, Fan L, et al. Overlap of burnout-depression symptoms among Chinese neurology graduate students in a national cross-sectional study. *BMC Med Educ* 2021;21(1):83.
PUBMED | CROSSREF

32. Kim RB, Park KS, Lee JH, Kim BJ, Chun JH. Factors related to depression symptom and the influence of depression symptom on self-rated health status, outpatient health service utilization and quality of life. *Korean J Health Educ Promot* 2011;28(1):81-92.

33. Seo JH, Kim HJ, Kim BJ, Lee SJ, Bae HO. Educational and relational stressors associated with burnout in Korean medical students. *Psychiatry Investig* 2015;12(4):451-8.
PUBMED | CROSSREF

34. Ryu S. Quality of life and quality of sleep in medical college students. *J Korean Soc Biol Ther Psychiatry* 2009;15(1):29-36.

35. Oginska H, Oginska-Bruchal K. Chronotype and personality factors of predisposition to seasonal affective disorder. *Chronobiol Int* 2014;31(4):523-31.
PUBMED | CROSSREF

36. Wolf MR, Rosenstock JB. Inadequate sleep and exercise associated with burnout and depression among medical students. *Acad Psychiatry* 2017;41(2):174-9.
PUBMED | CROSSREF

37. Grandner MA, Kripke DF, Yoon IY, Youngstedt SD. Criterion validity of the Pittsburgh Sleep Quality Index: investigation in a non-clinical sample. *Sleep Biol Rhythms* 2006;4(2):129-39.
PUBMED | CROSSREF

38. Lagnado DA, Sloman S. Learning causal structure. In: *Proceedings of the Annual Meeting of the Cognitive Science Society*; 2002 Aug 7–10; Fairfax, VA, USA. Seattle, WA, USA: Cognitive Science Society; 2002.

39. Morgan TL, McFadden T, Fortier MS, Tomasone JR, Sweet SN. Positive mental health and burnout in first to fourth year medical students. *Health Educ J* 2020;79(8):948-62.
CROSSREF

40. Cecil J, McHale C, Hart J, Laidlaw A. Behaviour and burnout in medical students. *Med Educ Online* 2014;19(1):25209.
PUBMED | CROSSREF

41. Park H, Lee YM, Ho MJ, Han HC. How the coronavirus disease 2019 pandemic changed medical education and deans' perspectives in Korean medical schools. *Korean J Med Educ* 2021;33(2):65-74.
PUBMED | CROSSREF