Depression Progression among Medical Students: A Longitudinal Analysis

Jiayu Yang, Shefali Shrivastava, Vishwanath Prathikanti

Introduction

Medical students are known for having disproportionately high mental health problems, with higher rates than others in the general population of depression and anxiety and burnout (Nair et al., 2023). Some of the major factors brought out previously that contributed a great deal to the students' mental well-being throughout their study included coping strategies, social support, and satisfaction with health (Cummerow et al., 2023). Although some longitudinal studies have been conducted, the trajectory of mental health issues throughout the entire medical school curriculum is not well understood (Schneider et al., 2023). Furthermore, past research documenting changes in metrics of mental health-such as depression and anxiety-exhibits inconsistent trends, with some studies showing increased symptoms over time (Rotenstein et al., 2016) while others report fluctuations (Puthran et al., 2016).

In this project, we shall examine data provided by the original longitudinal research of Carrard et al. (2024) into the mental health of medical students through a more detailed analysis to replicate findings of that study using regression analysis. Such an analysis of depression, anxiety, stress, burnout, and their biopsychosocial covariates will provide further insight into how such components do combine and interact to affect mental health outcomes. We also propose a Markov chain model representing the transition of states of mental health through time, hence showing in a dynamic perspective the course that students' mental health may take. It is with this dual approach-using both regression and Markov chain analyses-that we seek not only to confirm previous findings but also to add new light to the trends and fluctuations of the states of mental health as students progress through their medical studies.

Methods

Data Design and Collection

The data for this study was collected through an online questionnaire as part of the ETMED-L project, conducted in March 2021, November 2021, and November 2022 at the University of Lausanne, Switzerland (Carrard et al., 2024).

Key mental health variables include depression symptoms (CES-D), anxiety symptoms (STAI), and burnout indicators (MBI-SS). Biopsychosocial covariates include gender identity, social support, physical activity, health satisfaction, and coping strategies. A detailed summary of the descriptive statistics for these variables is provided in Appendix Tables A1–A3.

During data processing, researchers excluded 74 questionnaires that failed attention checks and 6 due to technical issues. Additionally, 168 students who repeated a year were removed to maintain a consistent curriculum trajectory for analysis. This yielded a final dataset with 1,595 unique students and 2,601 completed questionnaires: 909 students participated in one wave, 366 in two waves, and 320 in all three waves.

Our primary focus is on the students who participated in all three waves to facilitate longitudinal analysis of mental health trends and model changes over time. So in preparing the dataset, we structured it by retaining only their questionnaires and added columns for student_id and wave_number to facilitate tracking individual participants and their participation wave. This resulted in a student_id column ranging from 1 to 320 and a wave_number column with values 1, 2, and 3.

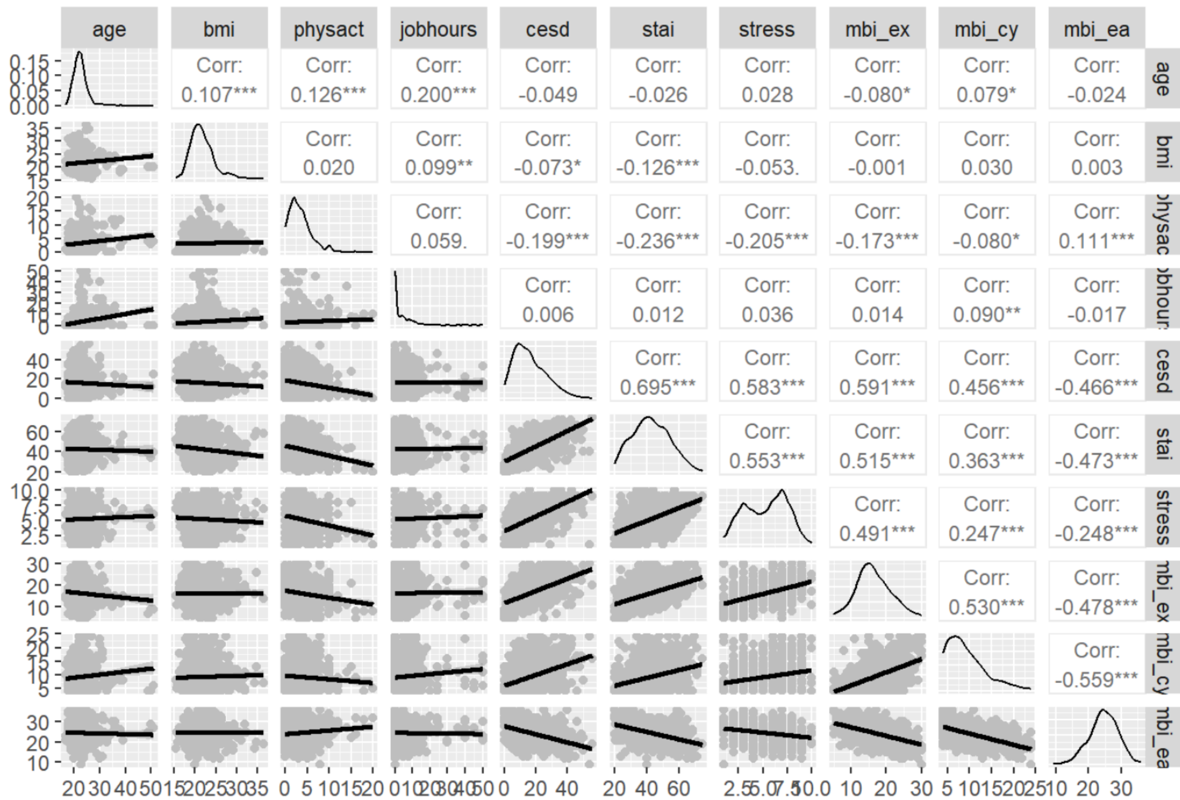
Then, two missing values were identified: age in row 814 and BMI in row 134. We imputed age as 23 (since participant 172 recorded as 22 in the prior year) and calculated an average BMI from participant 45 other responses to fill in the missing BMI.

Exploratory Data Analysis

The correlation analysis shows significant relations between various variables of mental health, burnout, and lifestyles. It has been noticed that symptoms of anxiety, STAI, and stress are strongly positively correlated with symptoms of depression, CES-D (Corr: 0.695 and 0.583, respectively, both $p < 0.001$), indicating that the higher the level of anxiety and stress, the higher the symptoms of depression. Among the components of burnout, there is a positive relation of emotional exhaustion-MBI-EX and cynicism-MBI-CY with symptoms of depression. Respective correlation coefficients are 0.591 and 0.456, $p < 0.001$, confirming that burnout may be related to higher depressive symptoms. Physical activity is inversely associated with depression, with a weak yet significant correlation. This may imply that physical activity could act as a protective agent. It also strongly reflects a minor positive relation to depressive symptoms concerning job hours. Academic self-efficacy is inversely related to the symptoms of depression; the correlation is -0.466, and the p-value is < 0.001. A low level of perceived academic self-efficacy would be related to a high level of depression. These results indicate a complex interaction of mental health with both burnout and lifestyle factors, underlining the potential influence anxiety, stress, and burnout might have on depression while pointing to the protective effect of regular physical activity.

Figure 1

Pairwise Scatterplots with Trend Lines

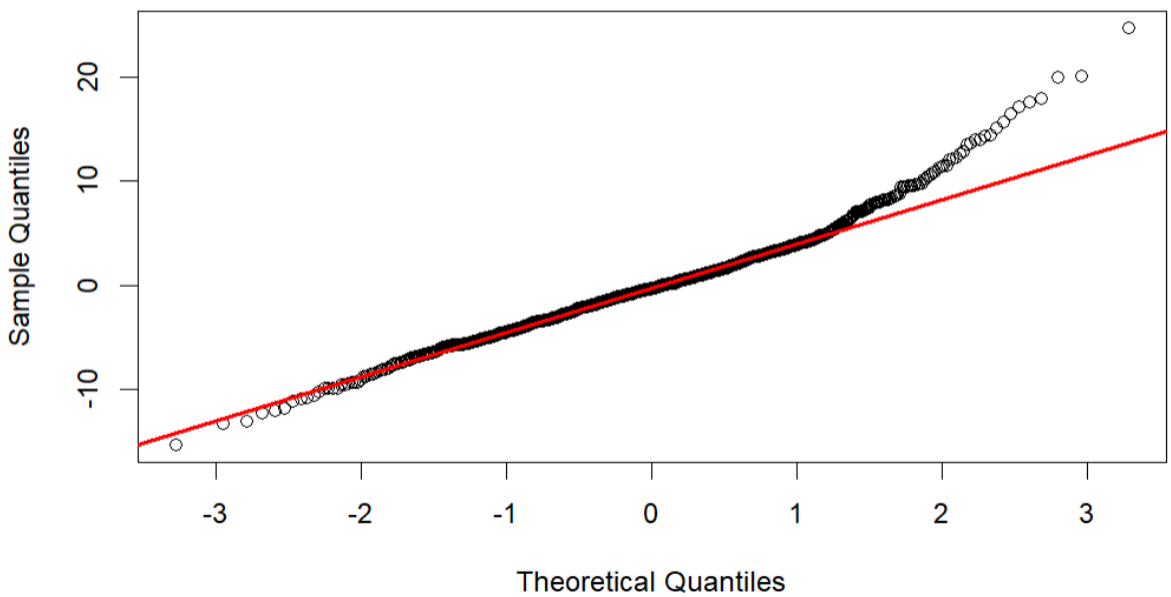


Assumptions

In our analysis, we checked the assumptions of the regression model through the statistical tests. Linearity was checked based on the correlations among \`cesd\` and the respective predictors; these were moderate in size for \`stai\`, \`stress\`, \`mbi_ex\`, and \`bdi_su\`, while lower for others such as \`year\` and \`educ_par\`. The multicollinearity assumption was evaluated using VIFs. All of the VIFs were less than 3, indicating that multicollinearity is not a serious problem. The Shapiro-Wilk test of normality was highly significant, $p < 2.2e\text{-}16$, which suggests that \`cesd\` is not normally distributed. Furthermore, a QQ plot (Fig 2) demonstrates evidence of non-normality. The Breusch-Pagan test indicated heteroscedasticity, $p = 5.667e\text{-}14$, which suggests variance in residuals is not constant and may want to consider robust standard errors or transformations. Our standardized residuals plot (Fig 3) supports this.

Figure 2

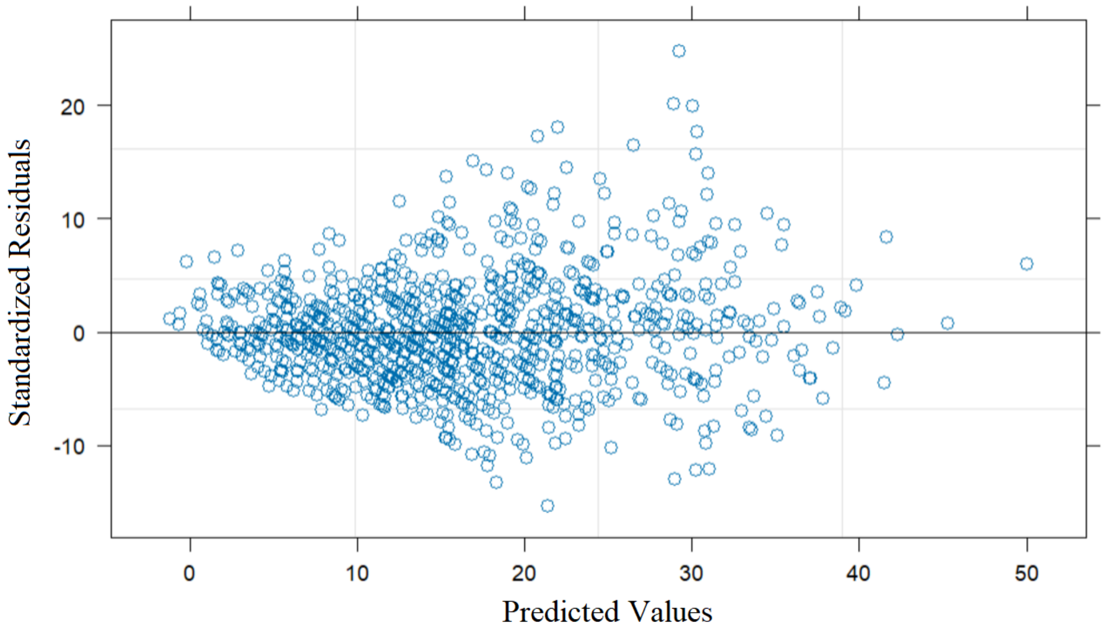
QQ Plot (Test of Normality)



Note: It is apparent from the QQ plot that the right tail shows signs of non-normality. While there may be a reasonable justification for normality, when used in conjunction with the Shapiro-Wilk test, we conclude the data is not normally distributed, and we will use HMM in our extended analysis to address this.

Figure 3

Standardized Residuals Plot: Residuals vs Fitted



Note: There appears to be an issue of heteroskedasticity. There is a visible funnel shape as the fitted values increase, where the residuals are more tightly clustered around 0 and spread out more widely as the fitted values grow larger.

The Durbin-Watson test reports an autocorrelation between the residuals, with a D-W statistic of 1.686574, and it may impact the inferences. If our data have a natural temporal structure, then maybe a model accounting for autocorrelation would be suitable. We saw that there are some concerns regarding normality and homoscedasticity and independence of residuals. With these assumptions not being satisfied, we will solve problems regarding normality, homoscedasticity, and residual independence so our results will appear reliable.

For our extended analysis, we applied the Gaussian hidden markov model to calculate hidden states based on CES-D scores of medical students. Our analysis assumes that the underlying depression states follow a first-order Markov process, where CES-D scores within each hidden state are normally distributed and conditionally independent. We presume time-homogeneous transition probabilities between states, with uniform measurement intervals. The model assumes finite, distinct depression severity levels with state-specific means and variances remaining constant over time, while measurement errors follow a Gaussian distribution.

Regression Model

In order to find the optimal regression model, we implemented stepwise regression to move from a full multilinear model to a more parsimonious one. When choosing between AIC vs BIC for model selection, we selected AIC. This is because our extension, Hidden Markov models, is an inherently predictive process, and due to the complexity and overlap of mental and physical health factors, the true model is likely complex.

We conducted a Variance Inflation Factor (VIF) analysis, and tabulated the results below. Because none of the variance inflation factor results were higher than a threshold of 3, we felt confident the data does not show evidence of multicollinearity.

Table 1

Variance inflation factors for each variable

Variable	Variance inflation factor
year	1.82
wave_number	1.64
health	1.15

psyt	1.12
cop_e	1.53
cop_h	1.33
fmale	1.18
socsup	1.36
educ_par	1.03
bdi_su	1.51
stai	2.33
stress	1.65
mbi_ex	1.97
mbi_cy	1.72

Note: As shown, no VIF exceeds 3, meaning we do not believe there is sufficient evidence of multicollinearity.

Hidden Markov Models

Markov chains model the behavior of a sequence of random variables over time based on the Markov process assumptions. It relies on a strong assumption that the current state probability only depends on the previous state (the Markov property).

$$P(q_t = \alpha \mid q_{t\ldots(t-1)} = P(q_t = \alpha)|q_{t-1}) \quad (1)$$

where q_i is the state at t^{th}time period.

We used Hidden Markov Models (HMMs) to identify latent states in depression trajectories, extending traditional Markov chains to incorporate unobservable states that must be inferred from observable data. The temporal progression of depression was measured using the Center for Epidemiologic Studies Depression Scale (CES-D), a 20-item self-report measure scored on a 4-point Likert scale (0-3), with total scores ranging from 0 to 60 (higher scores indicating greater depressive symptomatology). The observed data sequence $X = \{x_1, x_2, ..., x_{\square}\}$ represents CES-D scores at each timepoint t ($t = \{1,2,3\}$).

A systematic model selection process for the Gaussian HMM explored state spaces ranging from 2 to 10 states, with multiple random initialization. Covariance structure selection was not impactful for our univariate data compared to multivariate applications, and we only used the default (or spherical) covariance type. For each configuration, we performed multiple random initializations (n = 10) to mitigate local optima in parameter estimation. Models were fitted using the Baum-Welch algorithm, a specialized implementation of the Expectation-Maximization (EM) algorithm, with a maximum of 600 iterations to ensure convergence. The training data consisted of scaled CES-D scores, accounting for varying sequence lengths (based on time) across participants.

Model selection was conducted using the Akaike Information Criterion (AIC), which optimally balances model fit and complexity through penalty terms for additional parameters. This methodological choice was informed by empirical evidence demonstrating AIC's superior performance over BIC in shorter univariate time series analyses (Costa & Angelis, 2010). For each model configuration, the best log-likelihood was retained across initializations, and the model minimizing the AIC value was selected as optimal. The emission probability density function of our final selected model's emission probability density function is expressed as:

$$f^{(t)}_{(y \mid u)} = f_{Y^{(t)} \mid U^{(t)}}(y \mid u) \tag{2}$$

where t $\in$ {1, ..., T} represents the time points in our longitudinal data, $u \in \{0, ..., 4\}$ denotes the hidden states after combining states 1 and 2, $Y^{(t)}$ is the observed CES-D score (scaled) at time t, and $U^{(t)}$ is the hidden depression severity state at time t. Given the continuous nature of CES-D scores, we assume Gaussian emissions where $Y^{(t)}$ follows a normal distribution $N(\mu_{u^{(t)}}, \sigma^2_{u^{(t)}})$ with state-specific means and variances.

For model simplicity, this analysis focused solely on CES-D score trajectories over time without incorporating covariates; future iterations can extend this framework to include relevant factors such as academic stress and demographic variables through mixed HMM approaches.

Results

Regression Analysis

The AIC stepwise regression results are below. Of the original 27 variables, only 14 were included in the final model, 13 if we do not include wave_number, which was not significant at $\alpha = .05$. Aside from the wave number, all of these variables make logical sense when considered in the context of affecting depression. The original findings stated depression levels went down as students got more accustomed to the demands of medical school, and our result for the year variable is consistent with that. Much of the results make logical sense; satisfaction with one's health and having a proper social support network is associated with decreased depression, and having higher stress, emotional exhaustion and burnout is associated with higher levels of depression. Notable variables will be covered in the discussion section.

Table 2

Linear Regression Results

Variable	Coefficient	Standard Error	Test statistic	p-value
(Intercept)	0.21	1.81	0.11	0.908
year	-0.81	0.19	-4.31	<0.001***
wave_number	0.56	0.29	1.91	0.055
health	-0.74	0.21	-3.44	<0.001***
psyt	1.35	0.48	2.80	<0.01**
cop_e	0.15	0.06	2.30	<0.05*
cop_h	-0.22	0.08	-2.87	<0.01**
fmale	-1.12	0.43	-2.76	<0.01**
socsup	-0.46	0.12	-3.88	<0.001***
educ_par	0.53	0.22	2.39	<0.05*
bdi_su	2.80	0.25	11.00	<0.001***
stai	0.25	0.02	9.91	<0.001***
stress	1.02	0.11	9.06	<0.001***
mbi_ex	0.25	0.05	4.68	<0.01**
mbi_cy	0.23	0.06	4.15	<0.001***

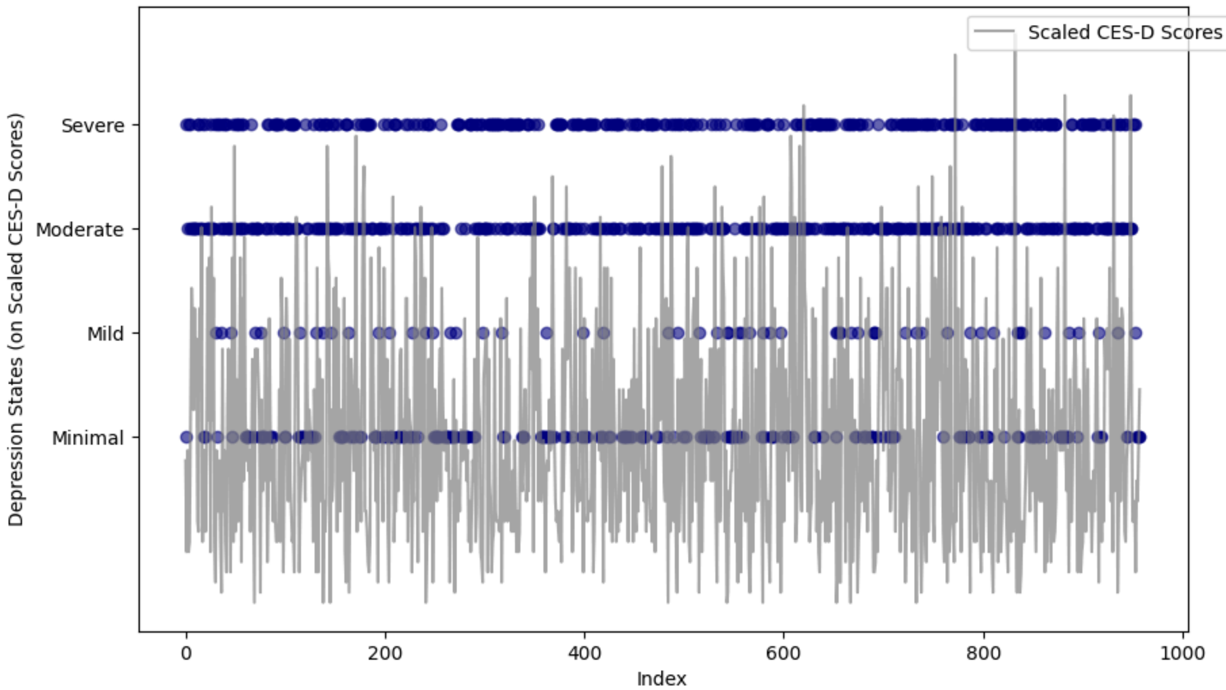
Note: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Extended Analysis: Hidden Markov Model

Hidden Markov Model analysis revealed four distinct states in the temporal analysis of scaled CES-D scores, ranging from minimal to severe symptom levels. The model categorized scaled depression scores into four stable states: State 4 (severe depression: highest sustained scores), State 3 (moderate depression), State 2 (mild depression), and State 1 (minimal depression). State predictions shown in Figure 4.

Figure 4

Gaussian HMM State Predictions (k = 4)



Additionally, the model's transition probabilities revealed distinct patterns in the progression of depressive symptoms among medical students (Figure 4). The model demonstrated highest stability in severe (0.40) and moderate (0.49) symptom states, suggesting these represent relatively stable clinical presentations. Mild and minimal severity states showed lower stability, with self-transition probabilities of 0.07 and 0.43 respectively, indicating these may be more transitory phases.

The transition patterns (Table 3) reveal several clinically relevant insights: First, the moderate depression state appears to act as a central "hub" with high incoming transition probabilities from all other states (0.29 from minimal, 0.45 from mild, and 0.48 from severe), suggesting it may represent a crucial phase for clinical intervention. Second, the progression from minimal to severe symptoms often occurs through intermediate states, but direct transitions to severe states are also possible (0.19), indicating that some students may experience rapid symptom escalation. Third, recovery patterns from severe states show a preference for transitioning to moderate states (0.48) rather than direct improvement to minimal symptoms (0.07), suggesting that recovery typically occurs gradually. These patterns highlight the importance of early intervention during moderate phases and continuous monitoring of students in minimal and mild states, given their potential for rapid progression to more severe states.

However, several limitations in the model fitting process warrant consideration. The optimal number of states showed sensitivity to random initializations, suggesting potential instability in the model structure. While the four-state solution emerged as optimal based on information criteria, alternative initializations produced varying state solutions, indicating that the underlying state structure may not be fully stable. This instability could reflect multiple factors: inherent complexity in the temporal patterns of depressive symptoms, limitations in the model's ability to consistently capture the true underlying state structure, constraints in the model design (particularly the univariate nature of the analysis), or limitations in the available data.

Table 3
Transition Probabilities Between Predicted Depression States in Medical Students

	Hidden Markov Model			
Depression State	1	2	3	4
1	0.43	0.09	0.29	0.19
2	0.31	0.07	0.45	0.17
3	0.2	0.04	0.49	0.28
4	0.07	0.05	0.48	0.4

Note. Model specifications: Univariate Gaussian HMM with spherical covariance (default), 4 states (n_components=4), 600 iterations (converged in 141). Model fit: Log likelihood = -1261.43, AIC = 2568.85, BIC = 2680.74.

Discussion

As the original paper discussed, the higher year someone is medical school, the lower their depression scores tend to be. Interestingly, however, age is not a significant factor in depression, meaning the low depression scores may be attributed to adapting to the medical school environment rather than maturing or other age-related causes.

Curiously, depression scores increase with having seen a psychotherapist, which may go against conventional wisdom. However, the question does not ask whether the student is in therapy sessions consistently, which may account for the observation.

Interestingly, emotion-based coping strategies ended up increasing depression levels, whereas help-based coping strategies decreased depression levels. Increasing awareness of proper resources for medical students or otherwise increasing the usage of help-based coping strategies may be a worthwhile strategy to combat depression.

Also worth noting is that identifying as a man decreases depression scores, necessitating more support for female and nonbinary medical students. This aligns with the original paper, that added Males tend to use different coping strategies and reported having better interactions with faculty.

Another key limitation of our approach lies in the underlying assumptions of the Hidden Markov Model. The model assumes Gaussian emissions and time-homogeneous transitions, which may not fully capture the complex, non-linear nature of depression trajectories. Additionally, our analysis relied solely on univariate CES-D scores, without incorporating other relevant mental health indicators or contextual factors that could influence state transitions. This univariate approach, while providing valuable insights into symptom progression patterns, may not adequately represent the multifaceted nature of depression in medical students. Future studies could benefit from incorporating multiple indicators and contextual variables to better understand the complex dynamics of mental health transitions in medical education. These methodological constraints, combined with the previously noted instability in state identification, suggest that while our findings provide important insights into depression trajectories, they should be interpreted within the context of these analytical limitations.

Reference:

Carrard, V., Berney, S., Bourquin, C., Ranjbar, S., Castelao, E., Schlegel, K., Gaume, J., Bart, P.-A., Marianne Schmid Mast, Preisig, M., & Berney, A. (2024). Mental health and burnout during medical school: Longitudinal evolution and covariates. *PloS One, 19*(4), e0295100–e0295100. https://doi.org/10.1371/journal.pone.0295100

Costa, M., & De Angelis, L. (2010). Model selection in hidden Markov models: a simulation study.

Cummerow, J., Obst, K., Voltmer, E., & Kötter, T. (2023). Medical students' coping with stress and its predictors: a cross-sectional study. *International Journal of Medical Education, 14*, 15–22. https://doi.org/10.5116/ijme.63de.3840

MacDonald, I.L., and Zucchini, W. (1997). Hidden Markov Models and Other Types of Models for Discrete-valued Time Series. London: Chapman & Hall.

Nair, M., Moss, N., Bashir, A., Garate, D., Thomas, D., Fu, S., Phu, D., & Pham, C. (2023). Mental health trends among medical students. *Proceedings (Baylor University. Medical Center), 36*(3), 408–410. https://doi.org/10.1080/08998280.2023.2187207

Puthran, R., Zhang, M. W. B., Tam, W. W., & Ho, R. C. (2016). Prevalence of depression amongst medical students: a meta-analysis. *Medical Education, 50*(4), 456–468. https://doi.org/10.1111/medu.12962

Rotenstein, L. S., Ramos, M. A., Torre, M., Segal, J. B., Peluso, M. J., Guille, C., Sen, S., & Mata, D. A. (2016). Prevalence of Depression, Depressive Symptoms, and Suicidal Ideation Among Medical Students. *JAMA, 316*(21), 2214. https://doi.org/10.1001/jama.2016.17324

Schneider, K., Breuer, G., Luibl, L., Paulsen, F., Scholz, M., & Burger, P. H. M. (2023). Vulnerable in the end – Longitudinal study among medical students on mental health and personal and work-related resources over a 5.5-year-period. *Annals of Anatomy - Anatomischer Anzeiger, 250*, 152155. https://doi.org/10.1016/j.aanat.2023.152155

Wiggins, L.M. (1973). Panel Analysis. Amsterdam: Elsevier.

Appendix:

Table A1

Descriptive Statistics for numeric predictors (N = 2601)

Variable	Description	N	Mean	SD	Min	Max
age	Age at time of questionnaire	2,601	22.28	3.14	17.00	51.00
bmi	Body Mass index	2,601	21.74	2.94	15.43	41.67
physact	Physical activity	2,601	3.24	2.65	0.00	21.00
cop_e	Emotion-focused coping usage	2,601	9.60	3.92	0.00	24.00
cop_p	Problem-focused coping usage	2,601	7.32	1.77	0.00	12.00
cop_h	Help-seeking coping usage	2,601	5.60	2.85	0.00	12.00
socsup	Social support score	2,601	8.08	1.92	0.00	10.00
jobhours	Hours of paid job per week	2,601	3.03	6.61	0.00	50.00
cesd	Depression symptoms on CES-D scale	2,601	18.26	11.20	0.00	58.00
bdi_su	Suicidal ideation (Beck depression inventory)	2,601	0.66	1.03	0.00	6.00
stai	Anxiety symptom score from State-Trait anxiety inventory	2,601	44.24	11.89	20.00	80.00
stress	Stress level (1-10 scale)	2,601	5.40	2.18	1.00	10.00

Variable	Description	N	Mean	SD	Min	Max
mbi_ex	Emotional exhaustion (Maslach burnout inventory)	2,601	16.64	5.03	5.00	30.00
mbi_cy	Cynicism (Maslach burnout inventory)	2,601	9.67	4.51	4.00	24.00
mbi_ea	Academic efficacy (Maslach burnout inventory, reversed dimension)	2,601	24.21	4.55	6.00	36.00

Note. Missing values were excluded from the current analysis but will be addressed in the next phase of the study.

Table A2

Descriptive Statistics for categorical variables (N = 2601)

Variable	Description	Levels	Count	Proportion
nwave	Number of survey waves participated in	1	909	34.95
		2	732	28.14
		3	960	36.91
longipart	Which wave participated in (1 = only first, 23 = first and second waves, 123 = all 3)	1	275	10.57
		2	240	9.23
		3	394	15.15

Variable	Description	Levels	Count	Proportion
		12	296	11.38
		13	60	2.31
		23	376	14.46
		123	960	36.91
sex	Gender identity 1 = male 2 = female 3 = nonbinary	1	813	31.26
		2	1,765	67.86
		3	23	0.88
Year	Curriculum year	1	597	22.95
		2	360	13.84
		3	467	17.95
		4	406	15.61
		5	426	16.38
		6	345	13.26
health	Satisfaction with health (1 = very unsatisfied, 5 = very satisfied)	1	75	2.88

Variable	Description	Levels	Count	Proportion
		2	244	9.38
		3	401	15.42
		4	1,319	50.71
		5	559	21.49
psyt	Consulted a psychotherapist during the past year (0 = no, 1 = yes)	0	1,977	76.01
		1	621	23.88
fmale	Identifying as male (1 = yes, 0 = no)	0	1,784	68.59
		1	817	31.41
part	Having partner (1 = yes, 0 = no)	0	1,200	46.14
		1	1,399	53.79
educ_par	Number of parents with a college or university degree	0	650	24.99
		1	556	21.38
		2	1,395	53.63

*Note***.** Missing values were excluded from the current analysis but will be addressed in the next phase of the study.

Table A3

Descriptive Statistics of categorical variables for CES-D Scores (N = 2601)

Variable		Levels	N	Mean	SD
nwave	Number of survey waves participated in	1	909	20.15760	11.842185
		2	732	18.52957	11.152654
		3	960	16.28646	10.272523
longipart	Which wave participated in (1 = only first, 23 = first and second waves, 123 = all 3)	1	275	18.29455	12.451239
		2	240	21.54622	11.948376
		3	394	20.62629	11.183322
		12	296	17.10169	11.168613
		13	60	17.05000	11.239194
		23	376	19.90054	10.981891
		123	960	16.28646	10.272523
sex	Gender identity 1 = male 2 = female 3 = nonbinary	1	813	14.77723	9.594111
		2	1,765	19.82356	11.516387

Variable		Level s	N	Mean	SD
		3	23	21.65217	12.085336
Year	Curriculum year	1	597	21.95439	11.615947
		2	360	19.91341	10.710648
		3	467	18.37554	11.219154
		4	406	16.52970	10.111125
		5	426	15.93176	10.710934
		6	345	14.95627	10.727396
health	Satisfaction with health (1 = very unsatisfied, 5 = very satisfied)	1	75	19.14667	15.761477
		2	244	27.88066	12.502486
		3	401	24.12469	10.575179
		4	1,319	17.42410	9.657264
		5	559	11.72043	8.527955
psyt	Consulted a psychotherapist during the past year (0 = no, 1 = yes)	0	1,977	16.48325	10.496751
		1	621	23.94175	11.502371
fmale	Identifying as male (1 = yes, 0 = no)	0	1,784	19.84009	11.526640

Variable		Levels	N	Mean	SD
		1	817	14.81773	9.607656
part	Having partner (1 = yes, 0 = no)	0	1,200	19.09137	11.228075
		1	1,399	17.55699	11.138885
educ_par	Number of parents with a college or university degree	0	650	18.59660	11.077388
		1	556	19.25451	11.749581
		2	1,395	17.71377	11.012959

Note. Missing values were excluded from the current analysis but will be addressed in the next phase of the study.