Longitudinal Regression Analysis of COVID-19 Epidemic Impacts on Adolescent Mental Health Outcomes

APH205 Final Report

Baiding Chen
Xi'an Jiaotong Liverpool University
AOP Biomedical statistics

1. Review of Epidemic Research conducted by Rong et al.

Methodological Limitations

Sampling and Generalizability

- **Single School Sample:** The participants were drawn exclusively from one junior high school in Jiangsu Province. This sample does not reflect the diversity of socioeconomic, cultural, and environmental factors influencing adolescent mental health across different geographical areas, thus limiting the external validity of the findings.
- **Homogeneity of Sample:** The participants are homogeneous in terms of age and educational stage, which may skew findings and limit generalizability to other age groups or educational contexts.

Data Collection and Instrumentation

- **Self-Report Bias:** The reliance on self-reported measures for mental health assessments introduces significant reporting biases, such as social desirability or recall bias. This could lead to misclassification of symptom severity and affects the accuracy of the results.
- **Lack of Clinical Validation:** Without clinical interviews or professional psychological assessments, the diagnoses based on self-reporting are not validated, potentially leading to overestimations or underestimations of mental health conditions.

Analytical Approach

- **Use of LLCA:** While longitudinal latent class analysis identifies distinct groups based on symptom trajectories, it assumes homogeneity within each group and stability over time. This can oversimplify the dynamic nature of mental health during the post-pandemic period and may result in overgeneralized interpretations.

Interpretative Concerns

Lack of Baseline Data on Psychological Status

- The absence of data on students' psychological status prior to and during the pandemic restricts the ability to assess the true impact of school closures and reopenings on mental health, limiting a comprehensive understanding of symptom trajectories.

Assumptions and Model Fit

- **Assumption of Negative Impacts:** The study's design is based on the assumption that the pandemic's effects are universally negative, disregarding any potential positive adaptations or resilience factors developed during the pandemic.

Timeliness and Relevance of Data

- **Delayed Data Usage:** The two-year lag in data analysis undermines the timeliness and applicability of the findings to current interventions or policies, possibly missing current trends or changes in mental health dynamics among adolescents.

2. Data Analysis on Epidemic

Data Pre-processing

We initiated our analysis with a thorough data cleaning process, addressing missing values, outliers, and inconsistencies within the dataset. Variables were re-coded and aligned with descriptions provided in Rong et al.'s study. In cases where necessary variables were missing from the dataset, we constructed these variables based on descriptions and other available data.

Descriptive Statistics

To understand the baseline impact of COVID-19 and the subsequent changes at the fourth follow-up, we engaged in descriptive statistical analysis. This involved calculating means, medians, and standard deviations, and visualizing the data distributions through semi-violin plots and box plots. Our semi-violin plots comparing the impact indices and PHQ mental health scores across different levels of COVID-19 impact on daily life showed a general trend of significant improvement in adolescents' mental health from the baseline to the follow-up period. Additionally, histograms of the frequency distribution of psychological test scores for GAD-7, ISI, and PHQ-9 demonstrated a pronounced shift towards lower severity scores at follow-up 4, indicating a marked decrease in depressive, anxiety, and insomnia symptoms among the students post-pandemic. These visualizations highlight the significant improvements in mental health as the impacts of the pandemic began to subside.

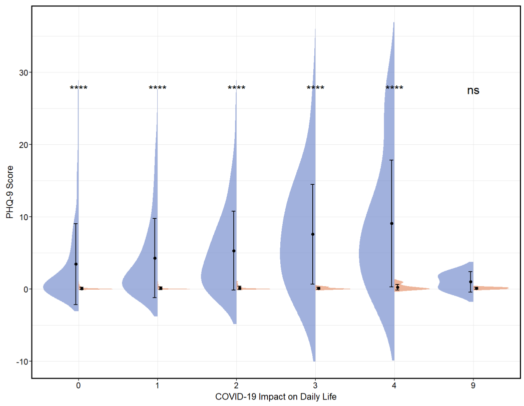

Figure 1: Semi-violin plots comparing indices and mental health scores

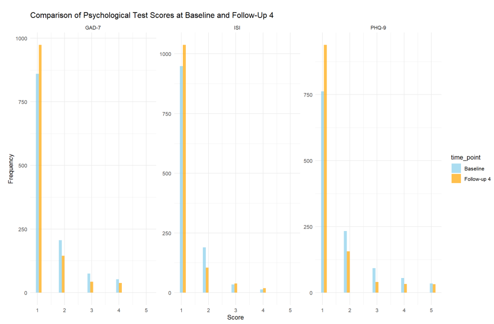

Figure 2: Histogram of the frequency distribution of psychological test scores

Exploratory Data Analysis

During the exploratory phase, we aimed to guarantee the pre-request of statistic method like Pearson. However, preliminary tests, including the Shapiro-Wilk test for normality and Q-Q plots, demonstrated that many of our variables did not conform to a normal distribution, prompting us to opt for Spearman's rank correlation instead of Pearson's. The histograms

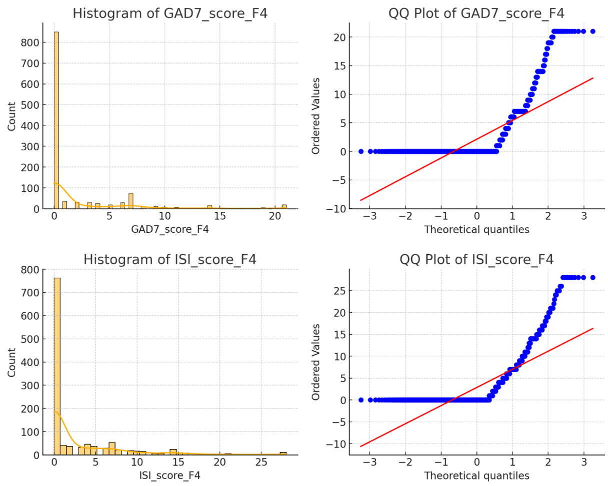

Figure 3: Histograms and Q-Q plots for the GAD-7 and ISI scores

and Q-Q plots for the GAD-7 and ISI scores at follow-up clearly illustrated the non-normal distribution, with the majority of scores clustering at the lower end and the data points in the Q-Q plots deviating significantly from the theoretical quantile line, especially at the tails. Therefore, we chose Spearman correlation for mining the relationship between data due to its robustness against non-normal distributions.

Model Development

Informed by our exploratory data analysis, we select significant variables and draw heat map which displayed the correlations between baseline perceptions of COVID-19 impacts and psychological symptom scores at the fourth follow-up. Significantly, the perceived impact on study practices (COVID19_impact_studypractice) showed a moderate positive correlation with scores of all three psychological conditions (depression, anxiety, and insomnia) assessed at the fourth follow-up. This suggests that greater negative impacts on study practices reported at baseline are associated with more severe psychological symptoms later. Additionally, the perceived impact on family income (COVID19_impact_familyincome) exhibited a moderate positive correlation with depression

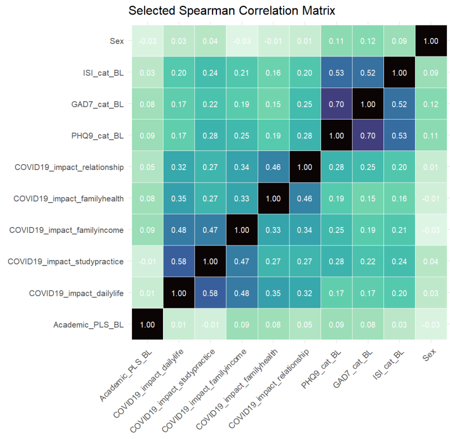

Figure 4: Heat map with the correlations between baseline and psychological symptom scores

and anxiety scores, indicating that financial strains linked to the pandemic might play a role in the exacerbation of these conditions.

Considering the mixed data types in our dataset—continuous, discrete, and categorical—we employed a Generalized Linear Model (GLM) to analyze these relationships. This model was particularly chosen for its flexibility in handling data which follow non-normal distributions and its ability to model various links between dependent and independent variables.

PHQ-9 Model Results

Model	term	estimate	std.error	statistic	p.value
PHQ-9	(Intercept)	-0.1512571	0.1257466	-1.2028726	0.2292702
PHQ-9	COVID19_impact_dailylife	-0.0000946	0.0072976	-0.0129675	0.9896559
PHQ-9	COVID19_impact_studypractice	0.0254211	0.0058004	4.3826048	0.0000128
PHQ-9	COVID19_impact_familyincome	0.0069990	0.0065004	1.0766963	0.2818390
PHQ-9	COVID19_impact_familyhealth	0.0116502	0.0081320	1.4326470	0.1522272
PHQ-9	COVID19_impact_relationship	0.0199183	0.0090757	2.1946883	0.0283830
PHQ-9	Sex	0.0432231	0.0115560	3.7403232	0.0001927
PHQ-9	Age_BL	0.0064646	0.0081616	0.7920744	0.4284787
PHQ-9	BMI	0.0018855	0.0016823	1.1208350	0.2625893
PHQ-9	Academic_Year	-0.0384095	0.0136458	-2.8147564	0.0049637
PHQ-9	Sibling_BL	-0.0142235	0.0099582	-1.4283192	0.1534683
PHQ-9	Academic_PLS_BL	0.0075478	0.0053893	1.4005183	0.1616247
PHQ-9	Morbidity_S_BL	0.0621736	0.0308926	2.0125713	0.0443897
PHQ-9	Morbidity_P_BL	-0.0072781	0.0225644	-0.3225478	0.7470957

Figure 5: GLM regression Result on PHQ-9 (all data available on appendix)

The regression results revealed that the perceived impact on study practices (COVID19_impact_studypractice) significantly increased PHQ-9 scores, with a coefficient of 0.0254 (p-value ¡ 0.0001), indicating a strong association between increased educational disruptions due to COVID-19 and elevated levels of depressive symptoms. Furthermore, the variable 'Sex' showed a coefficient of 0.0432 (p-value = 0.0019), suggesting that gender differences play a crucial role in the psychological impact experienced, with females likely experiencing higher levels of depressive symptoms than males during the pandemic. Additionally, the model coefficients comparison with confidence interval chart (Fig) across different psychological assessments (PHQ-9, GAD-7, and ISI) further illuminated the distinct effects of these predictors across various forms of mental health challenges.

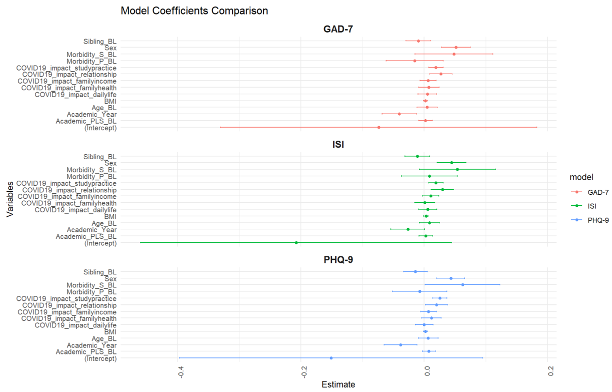

Figure 6: Confidence intervals of GLM on multiple levels

Stratified Analysis

Our findings from the GLM led us to conduct further analyses stratified by age and gender to explore differential impacts more specifically. For instance, displaying by density and violin plots, older adolescents (15-16) and females generally exhibited higher levels of psychological distress, as indicated by denser and higher peaks in these plots, suggesting heightened sensitivity to pandemic-related stressors.

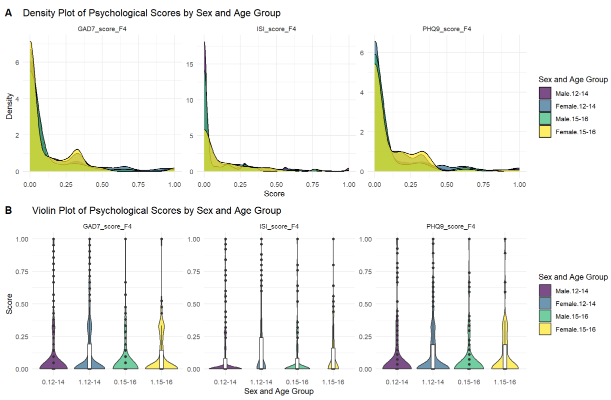

Figure 7: Descriptive analysis of stratified data by age and gender

Building upon these descriptive insights, we employed stratified GLM analysis to rigorously assess the relationships between COVID-19 impacts and psychological outcomes within these subgroups. The analysis revealed that certain factors, such as the perceived impact on study practice and family health, had notably stronger positive correlations with psychological distress in older adolescents and females.

A. Male

```
Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                  -0.013076   0.041470  -0.315 0.752632
COVID19_impact_dailylife      0.007669   0.009339   0.821 0.411872
COVID19_impact_studypractice  0.024115   0.007222   3.339 0.000892 **
COVID19_impact_familyincome   0.005797   0.008256   0.702 0.482861
COVID19_impact_familyhealth   0.027912   0.010243   2.725 0.006612 **
COVID19_impact_relationship   0.017302   0.011213   1.543 0.123337
BMI                           0.001763   0.002040   0.864 0.387748
```

B. Female

```
Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                   0.038795   0.057019   0.680  0.49654
COVID19_impact_dailylife     -0.013289   0.011585  -1.147  0.25184
COVID19_impact_studypractice  0.027046   0.009446   2.863  0.00435 **
COVID19_impact_familyincome   0.010560   0.010324   1.023  0.30679
COVID19_impact_familyhealth  -0.009097   0.013081  -0.695  0.48709
COVID19_impact_relationship   0.025057   0.015068   1.663  0.09689 .
BMI                           0.002427   0.002880   0.843  0.39975
```

C. Age 12-14

```
Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                  0.0135369  0.0387491   0.349  0.72691
COVID19_impact_dailylife     0.0001042  0.0085699   0.012  0.99030
COVID19_impact_studypractice 0.0268320  0.0066323   4.046 5.65e-05 ***
COVID19_impact_familyincome  0.0092206  0.0075607   1.220  0.22295
COVID19_impact_familyhealth  0.0052596  0.0098868   0.532  0.59487
COVID19_impact_relationship  0.0311455  0.0110816   2.811  0.00505 **
BMI                          0.0018635  0.0019230   0.969  0.33278
```

D. Age 15-16

```
Coefficients:
                               Estimate Std. Error t value Pr(>|t|)
(Intercept)                   0.0676436  0.0684896   0.988   0.3243
COVID19_impact_dailylife     -0.0017248  0.0135335  -0.127   0.8987
COVID19_impact_studypractice  0.0211736  0.0119757   1.768   0.0783 .
COVID19_impact_familyincome  -0.0083758  0.0123233  -0.680   0.4974
COVID19_impact_familyhealth   0.0349856  0.0136936   2.555   0.0112 *
COVID19_impact_relationship   0.0013963  0.0152867   0.091   0.9273
BMI                          -0.0007484  0.0034556  -0.217   0.8287
```

Figure 8: Stratified GLM analysis results

To validate the reliability of our models and ensure the robustness of our findings, we further analyzed the model coefficients across different groups and presented them along with their confidence intervals. The chart displaying model coefficients by group showcased how demographic variables modified the effects of COVID-19 impacts, with significant variations evident across different age and gender categories.

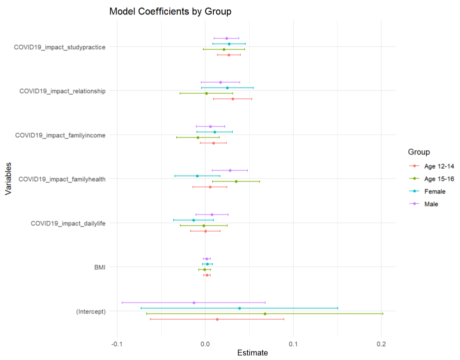

Figure 9: Confidence intervals of GLM on multiple levels

Main Text Words Count: 1011

3. Appendix

R code of data analysis

Due to the R language analysis code is too long (about 400 lines) will occupy a lot of space, in order to maintain the simplicity and beauty of this coursework, all with this paper data analysis related to the R language code as well as high-definition full-size original graphs are stored in the github for professors to view! URL: github.com/shellwork/APH205_analysis_Rcode

All heat map results

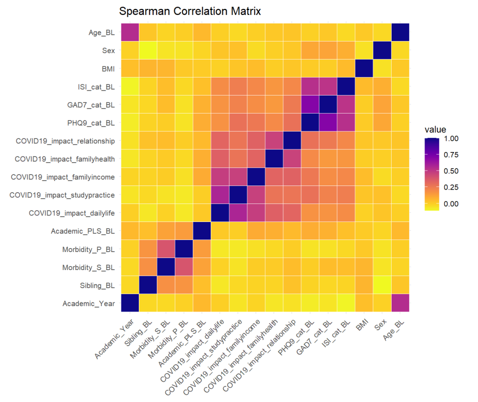

Figure 10: Whole heat map results

All GLM regression results

Model	term	estimate	std.error	statistic	p.valu
		Model Results			
PHQ-9	(Intercept)	-0.1512571	0.1257466	-1.2028726	0.22927
PHQ-9	COVID19_impact_dailylife	-0.0000946	0.0072976	-0.0129675	0.98965
PHQ-9	COVID19_impact_studypractice	0.0254211	0.0058004	4.3826048	0.00001
PHQ-9	COVID19_impact_familyincome	0.0069990	0.0065004	1.0766963	0.28183
PHQ-9	COVID19_impact_familyhealth	0.0116502	0.0081320	1.4326470	0.15222
PHQ-9	COVID19_impact_relationship	0.0199183	0.0090757	2.1946883	0.02838
PHQ-9	Sex	0.0432231	0.0115560	3.7403232	0.00019
PHQ-9	Age_BL	0.0064646	0.0081616	0.7920744	0.42847
PHQ-9	BMI	0.0018855	0.0016823	1.1208350	0.26258
PHQ-9	Academic_Year	-0.0384095	0.0136458	-2.8147564	0.00496
PHQ-9	Sibling_BL	-0.0142235	0.0099582	-1.4283192	0.15346
PHQ-9	Academic_PLS_BL	0.0075478	0.0053893	1.4005183	0.16162
PHQ-9	Morbidity_S_BL	0.0621736	0.0308926	2.0125713	0.04438
PHQ-9	Morbidity_P_BL	-0.0072781	0.0225644	-0.3225478	0.74709
GAD-7	(Intercept)	-0.0740717	0.1309759	-0.5655368	0.57181
GAD-7	COVID19_impact_dailylife	0.0050969	0.0076011	0.6705472	0.50264
GAD-7	COVID19_impact_studypractice	0.0190041	0.0060417	3.1455080	0.00170
GAD-7	COVID19_impact_familyincome	0.0062597	0.0067708	0.9245238	0.35540
GAD-7	COVID19_impact_familyhealth	0.0076499	0.0084701	0.9031595	0.36662
GAD-7	COVID19_impact_relationship	0.0271094	0.0094531	2.8677852	0.00420
GAD-7	Sex	0.0514910	0.0120365	4.2778879	0.00002
GAD-7	Age_BL	0.0049267	0.0085010	0.5795360	0.56233
GAD-7	BMI	0.0019446	0.0017522	1.1097884	0.26731
GAD-7	Academic_Year	-0.0405104	0.0142132	-2.8501871	0.00444
GAD-7	Sibling_BL	-0.0094406	0.0103724	-0.9101724	0.36292
GAD-7	Academic_PLS_BL	0.0021193	0.0056134	0.3775379	0.70584
GAD-7	Morbidity_S_BL	0.0481147	0.0321773	1.4952998	0.13510
GAD-7	Morbidity_P_BL	-0.0154972	0.0235028	-0.6593767	0.50978
ISI	(Intercept)	-0.2079931	0.1289235	-1.6133070	0.10694
ISI	COVID19_impact_dailylife	0.0057412	0.0074820	0.7673292	0.44304
ISI	COVID19_impact_studypractice	0.0189883	0.0059470	3.1929169	0.00144
ISI	COVID19_impact_familyincome	0.0108109	0.0066647	1.6221161	0.10504
ISI	COVID19_impact_familyhealth	0.0010140	0.0083374	0.1216180	0.90322
ISI	COVID19_impact_relationship	0.0296639	0.0093050	3.1879644	0.00147
ISI	Sex	0.0444930	0.0118479	3.7553442	0.00018
ISI	Age_BL	0.0084576	0.0083678	1.0107245	0.31235
ISI	BMI	0.0028202	0.0017248	1.6351544	0.10228
ISI	Academic_Year	-0.0265647	0.0139905	-1.8987621	0.05784
ISI	Sibling_BL	-0.0112131	0.0102098	-1.0982702	0.27231
ISI	Academic_PLS_BL	0.0023631	0.0055254	0.4276762	0.66896
ISI	Morbidity_S_BL	0.0539616	0.0316731	1.7037049	0.08870
ISI	Morbidity_P_BL	0.0084595	0.0231345	0.3656678	0.71467

Figure 11: Whole GLM regression analysis results