Identifying heterogeneous treatment effects for online single-session interventions for adolescent depression: a secondary analysis

Word count: 2998

Abstract

Background: Two online single-session interventions (SSIs), Project ABC and Project Personality, have demonstrated effectiveness in treating adolescent depression in a randomized clinical trial. However, the heterogeneity of treatment effect (HTE) remains unclear. A better understanding of HTE could facilitate individualized clinical decision-making.

Objectives: To evaluate the HTE of Project ABC and Project Personality for adolescent depression.

Methods: A subset of participants from a previous randomized trial was analyzed to study the HTE of Project ABC and Project Personality. Two study populations were formed: the first comprised participants assigned to either Project ABC or Placebo control, and the second included participants assigned to either Project Personality or Placebo Control. A baseline depression severity prediction model was constructed for each SSI using Elastic Net under a leave-one-out cross-validation framework. HTE was assessed using a risk-based approach by including the statistical interaction between treatment assignment and linear predictor of baseline depression severity score in the multivariable linear regression. Treatment effects were assessed across decile groups based on baseline depression severity scores.

Findings: Baseline depression severity varies among participants with different clinical characteristics. Individuals with a non-binary gender identity, those who self-identified as LGBTQ+, and those who do not use any coping strategies are more likely to develop severe depression and derive less benefit from the interventions. The average treatment effects (ATE) were -0.080 (s.d.: 0.022) for Project ABC and -0.069 (s.d.: 0.023) for Project Personality. Heterogeneous treatment effects (HTE) were statistically significant for Project ABC but not for Project Personality. "Risk" stratification results indicated HTE for both interventions. Participants in the lowest depression score decile benefited the most from the online SSIs, with a mean difference in the 3-month CDI-SF mean score of -0.16 (bootstrapped CI: [-0.20, -0.04]) for Project ABC and -0.10 (bootstrapped CI: [-0.15, 0.01]) for Project Personality. Both SSIs were not beneficial for participants in the two highest score deciles.

Interpretations: This study identified the varying effectiveness of online SSIs for reducing depression symptoms across adolescents. Individual clinical characteristics should be considered when assigning these interventions.

1   Introduction

The COVID-19 pandemic has significantly impacted the mental health of youth, limiting their access to effective depression treatments. Online single-session interventions (SSIs) have emerged as a promising solution, offering accessible, brief, and targeted interventions for adolescent depression. During the pandemic, the adoption of SSIs has increased, and their utility for depression treatment has been examined in several randomized trials, demonstrating their potential efficacy [1, 2].

One randomized trial conducted by Schleider et al. [3] investigated the effectiveness of two self-guided online SSIs, Project ABC [3] and Project Personality [1, 2], versus an active control for U.S. adolescents aged 13 to 16 during three months of the COVID-19 pandemic. The results showed that, after adjusting for baseline depression, both Project ABC and Project Personality reduced 3-month depressive symptoms compared to the active control group. However, the authors only reported the average treatment effect (ATE), a measure of the average effect of a treatment across all study participants[4]. The HTE of Project ABC and Project Personality has not been studied yet. In practice, the treatment effect can differ across individuals or subgroups with different clinical characteristics. This phenomenon is referred to as the heterogeneity of treatment effect (HTE), defined as nonrandom differences in the direction or magnitude of a treatment's clinical effect across different levels of patient characteristics [5].

This study is a secondary analysis of the randomized trial by Schleider et al.[3]. The objective of this study is to identify subgroups of participants with heterogeneous effects for each online SSI. Specifically, this study aims to build two separate models to investigate the HTE for Project ABC SSI versus the control group, and for Project Personality SSI versus the control group, utilizing baseline demographic and clinical predictors. These models aim to enhance the understanding of how different baseline characteristics influence the efficacy of each intervention, thereby informing more tailored strategies for treatment assignment for depression symptoms.

2   Methodology

2.1   Study participants and variables

This study utilized a subset of participants from the original randomized trial, where participants completed a baseline survey and a 3-month follow-up survey between November 2020 and March 2021. The sample consists of U.S. adolescents aged 13 to 16 who exhibited elevated depressive symptoms, indicated by a score of ≥ 2 on the Patient Health Questionnaire-2 (PHQ-2) [6].

Baseline characteristics were collected through the baseline survey and included clinical characteristics such as biological sex, age, race/ethnicity, sexual orientation, family challenges during the COVID-19 pandemic, coping strategies to improve mental health during the COVID-19 pandemic, gender identity, and , baseline depression severity ,quantified by the Children's Depression Inventory-Short Form (CDI-SF) score [7].

The interventions studied were one of the following: Project ABC, Project Personality and Placebo control. The primary outcome was the 3-month CDI-SF mean score following the intervention. Detailed descriptions of all variables can be found in the variable dictionary in Appendix Table 4.

2.2   Statistical analysis

2.2.1   Variable recoding

Some of the originally collected variables could not be directly used for modeling. Collapsing categories strategy and latent class analysis (LCA) [8] were used to recode these variables.

First, variables that had multiple categories with limited sample sizes in some strata were collapsed. Specifically, sexual orientation, family challenges, and race were reclassified.

A second strategy, LCA, was applied to the gender identity and coping strategies variables, as these variables consist of multiple question items that cannot be easily classified into mutually exclusive categories. For instance, the gender identity question includes 16 items, resulting in $2^{16} = 65,536$ possible subgroups. Treating each item as a binary variable is impractical, as it would significantly reduce the statistical power needed to detect potential treatment effects. Additionally, applying penalized methods like Lasso is unsuitable because these items are not independent and may collectively reflect intrinsic patterns of unobserved variables. Omitting some items could result in a loss of valuable information. Since gender identity and coping strategies are social behavior variables, it is plausible that the observed responses are influenced by underlying, unobserved variables. These considerations led to the use of LCA, which helps identify underlying subgroups of individuals based on their responses to observed items.

The application of LCA followed published guidelines [9]. Latent models with increasing numbers of latent classes (two, three, and so on) were constructed, and model fit was examined using the G^2 statistic. Model selection was based on information criteria (AIC, BIC) and the interpretability of latent classes. The selected latent model was used to estimate the probability of each individual belonging to each latent class. Individuals were then assigned to the latent class for which they had the highest probability of membership. This latent class variable was then used in subsequent analyses. The LCA was conducted using the `poLCA` (version 1.6.0.1) R package [10].

2.2.2 Risk-based Approach for HTE

HTE was explored using a risk-based approach [11, 12]. First, a baseline depression severity prediction model was constructed to capture the variation in depression severity attributable to individual characteristics other than treatment. The dataset was randomly split into a derivation set (80%) and a test set (20%).

A prediction model was developed on the derivation set using multivariable linear regression, where the primary outcome (3-month follow-up CDI-SF mean score) was regressed on baseline characteristics including race, biological sex, sexual orientation, gender identity, type of family challenges and type of coping strategies. To obtain an honest estimate, a leave-one-out cross-validation (LOOCV) framework was employed. Both ℓ_1 and ℓ_2 penalties were considered for variable selection, resulting in an Elastic Net model. The hyperparameters that resulted in the lowest mean squared error were chosen to build the final baseline depression severity prediction model on the entire sample. The calibration slope and calibration-at-large of the baseline prediction model were evaluated on the validation and test sets, respectively. The baseline prediction model was built using the `caret` (version 6.0-94) R package [13].

Next, the 3-month CDI-SF mean score was predicted using the constructed baseline prediction model, and predicted score was evaluated as a potential effect modifier. HTE was assessed using a multivariable linear model with four terms: baseline CDI-SF mean score, predicted baseline depression severity score, a treatment indicator variable, and an interaction term between treatment and predicted baseline depression severity score. After adjusting for the baseline CDI-SF mean score, the treatment indicator coefficient measured the extent to which there was a constant change in the 3-month depression score. The interaction term determined whether there was a differential treatment benefit across different subgroups.

For comparison, a main effect model was fitted to estimate the average treatment effect (ATE) by regressing the 3-month CDI-SF mean score on the baseline CDI-SF mean score and treatment condition. HTE were assessed by testing the significance of the interaction term and comparing the mean difference in the 3-month CDI-SF mean score across individuals. Specifically, participants were divided into 10

groups based on deciles of their predicted 3-month CDI-SF mean scores, forming 10 baseline "risk" subgroups. The conditional average treatment effect (cATE) was then computed within each decile of these score groups. These cATE values were compared across score groups and with the ATE. Bootstrap methods were used to compute the confidence intervals for HTE across different score subgroups.

The study samples were divided into two datasets: one group comprising participants from Project ABC and the Placebo Control, and the second group consisting of participants from Project Personality and the Placebo control. The baseline depression severity score and HTE models were constructed separately for each of these two groups. The HTEs for Project ABC and Project Personality were evaluated in two datasets independently.

2.2.3 Sensitivity analysis

Missing data was handled following the guidelines in Mirzaei et al. (2022) [14]. Complete case analysis (CCA) was implemented as the missingness rate was below 5%. An inverse probability of missingness weighting (IPMW) analysis [15] was conducted to impute missing data, where the missing status was regressed on fully observed baseline variables. The HTE effects were assessed using the same strategy as in CCA, and the results were compared with CCA. Results can be found in Appendix **??**.
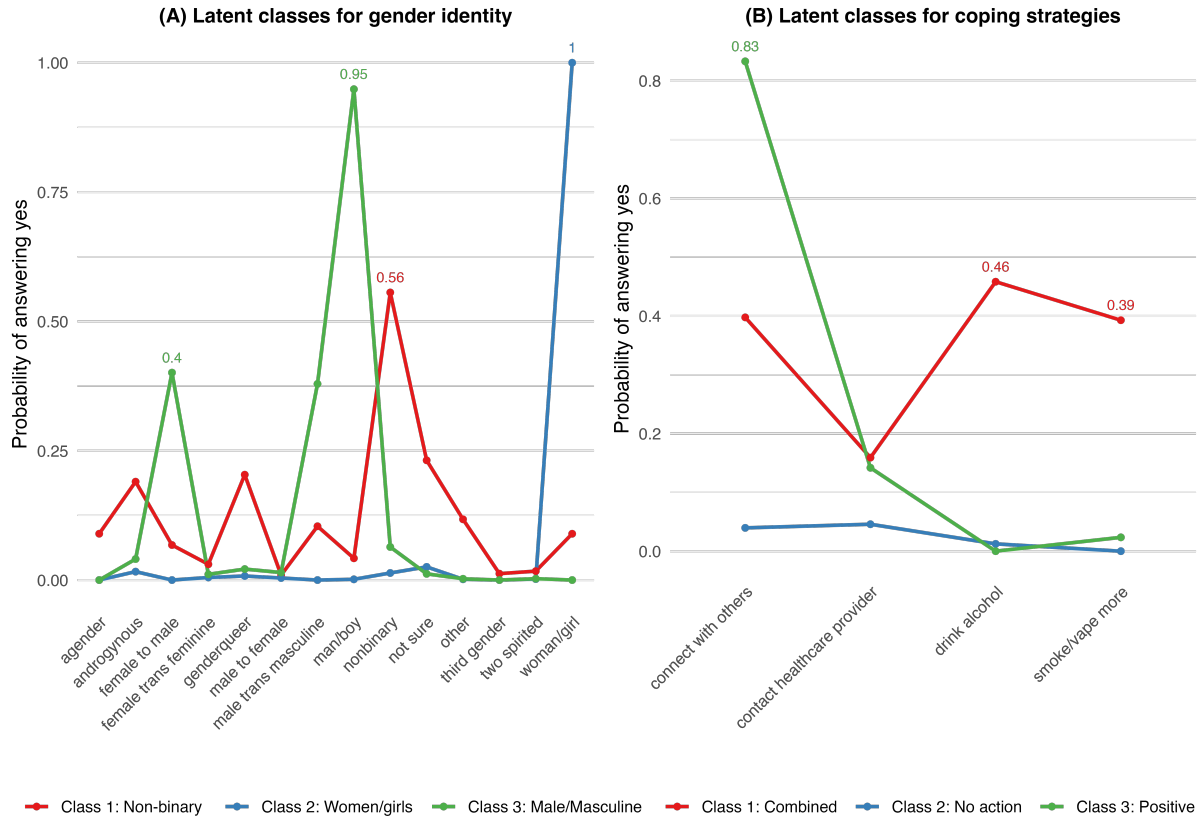
All analysis was conducted with R (version 4.3.1)[16]. The codes can be accessed at this github repository.

3 Results

3.1 Participant characteristics

Sexual orientation was reclassified into three categories: "Heterosexual", "LGBTQ+" (encompassing individuals who self-identified as Gay/Lesbian/Homosexual, Bisexual/Pansexual, Queer, or Unsure/Questioning), and "Other" (including individuals who self-identified as Other, Asexual, or did not use a specific label). Family challenge was condensed into three categories: "No Challenge in the Past Two Weeks", "School-only Challenge", and "Other" (including financial, other, or mixed challenges). For race, participants who selected more than one race were classified as "Mixed".

The LCA approach classified individuals into different latent classes for gender identity and coping strategies, as shown in Figure 1. This figure displays the probability of answering "yes" to survey questions by latent class and illustrates the response patterns of each class. Figure 1A identified three distinct response patterns/subgroups for gender identity: "Non-binary", "Women/Girls", and "Male/-Masculine". For coping strategies, LCA indicated a latent variable with three classes: "Positive Coping", "Combined Coping", and "No Coping" (Figure 1B).

(A) Latent classes for gender identity

(B) Latent classes for coping strategies

Class 1: Non-binary • Class 2: Women/girls • Class 3: Male/Masculine • Class 1: Combined • Class 2: No action • Class 3: Positive

Figure 1: Probability of answering "yes" to each item of gender identity and cooping strategies questions (A) LCA results for gender identity; (B) LCA results for coping strategies

A total of 1,441 adolescents were included in the analysis after removing participants with incomplete data. There were 489 (34%), 464 (32%), and 488 (34%) participants in the Project ABC, Project Personality, and Placebo control groups, respectively. Baseline clinical characteristics by treatment condition are presented in Table 1. The average age of participants was 15 years (s.d.: 0.9), with 1,289 (90%) being female and 152 (10%) being male. Over half (798, 55%) of the participants were white. The most common COVID-19 challenge was school-related, accounting for 55% to 60%, and around half of the participants did not take any action to improve their mental health.

Table 1: Baseline characteristics of 1441 participants from the online single-session interventions for adolescent depression randomized trial

| | Treatment received | | |
| | Project ABC | Project Personality | Placebo Control |
Demographics	**N = 489**	**N = 464**	**N = 488**
Baseline CDI-SF[1] mean score (0-2)	1.15 (0.34)	1.17 (0.36)	1.16 (0.35)
Race/ethnicity			
White	262 (54%)	262 (56%)	274 (56%)
Asian Including Asian Desi	58 (12%)	50 (11%)	50 (10%)
Black/African-American	40 (8.2%)	36 (7.8%)	33 (6.8%)
Hispanic/Latinx	61 (12%)	53 (11%)	57 (12%)
Mixed[2]	68 (14%)	63 (14%)	74 (15%)
Age (yrs)	15.15(0.92)	15.20 (0.89)	15.20 (0.90)

Age (yrs)			
13	32 (6.5%)	28 (6.0%)	28 (5.7%)
14	81 (17%)	63 (14%)	77 (16%)
15	156 (32%)	162 (35%)	150 (31%)
16	220 (45%)	211 (45%)	233 (48%)
Biological sex			
Female	437 (89%)	418 (90%)	434 (89%)
Male	52 (11%)	46 (9.9%)	54 (11%)
Language			
English	476 (97%)	450 (97%)	476 (98%)
Other	13 (2.7%)	14 (3.0%)	12 (2.5%)
Sexual orientation[3]			
Heterosexual	97 (20%)	106 (23%)	108 (22%)
LGBTQ+	327 (67%)	291 (63%)	309 (63%)
Other	65 (13%)	67 (14%)	71 (15%)
Type of challenges[4]			
School	293 (60%)	255 (55%)	271 (56%)
Other	97 (20%)	124 (27%)	119 (24%)
No challenges	99 (20%)	85 (18%)	98 (20%)
Total number of challenges			
0	99 (20%)	85 (18%)	98 (20%)
1	297 (61%)	275 (59%)	276 (57%)
>=2	93 (19%)	104 (22%)	114 (23%)
Gender identity[5]			
Non-binary	98 (20%)	86 (19%)	102 (21%)
Women/girls	325 (66%)	306 (66%)	313 (64%)
Male/Masculine	66 (13%)	72 (16%)	73 (15%)
Type of coping strategies[6]			
Positive	193 (39%)	175 (38%)	202 (41%)
Combined	61 (12%)	53 (11%)	58 (12%)
No action	235 (48%)	236 (51%)	228 (47%)

All categorical variables are presented as N (%); all continuous variables are presented as mean (s.d.); Percentages may not sum up to 100% due to rounding errors.

[1] Children's Depression Inventory 2 Short-Form;

[2] Mixed race includes individuals who selected more than one racial category;

[3] LGBTQ+ encompasses individuals who self-identified as one of the following: Gay/Lesbian/Homosexual, Bisexual/Pansexual, Queer, or Unsure/Questioning. Other refers to sexual orientations not listed in the original survey;

[4] Specific challenges during the COVID-19 pandemic. School refers to not being able to attend school in person or at all. Other includes financial or other challenges not listed in the original survey

[5] Latent variable for gender identity. Gender identity was classified into three latent classes based on the latent class analysis results: Non-binary, Women/Girls, and Male/Masculine;

[6] Latent variable for coping strategies to improve mental health during the COVID-19 pandemic. Positive coping strategies include connecting with others/contacting healthcare workers. Combined refers to individuals who use both positive and negative coping strategies (e.g., drinking alcohol/smoking or vaping more) at the same time. No action refers to individuals who did not take any actions.

3.2 Baseline Depression Severity Score Prediction Model

A total of 977 subjects were used to construct the baseline depression severity prediction model for Project ABC and control samples, and 952 subjects were used for Project Personality and control samples. The model results are presented in Table 2. The final predictors included in the baseline prediction models for both samples were: gender identity, race, biological sex, sexual orientation, type of challenges, and coping strategies. The likelihood ratio test results showed that gender identity, race, biological sex, sexual orientation, and coping strategies were statistically associated with depression in the Project ABC and control samples. For Project Personality and control population, gender identity and coping strategies were associated with depression. A detailed likelihood ratio test results are provided in Appendix Table 5.
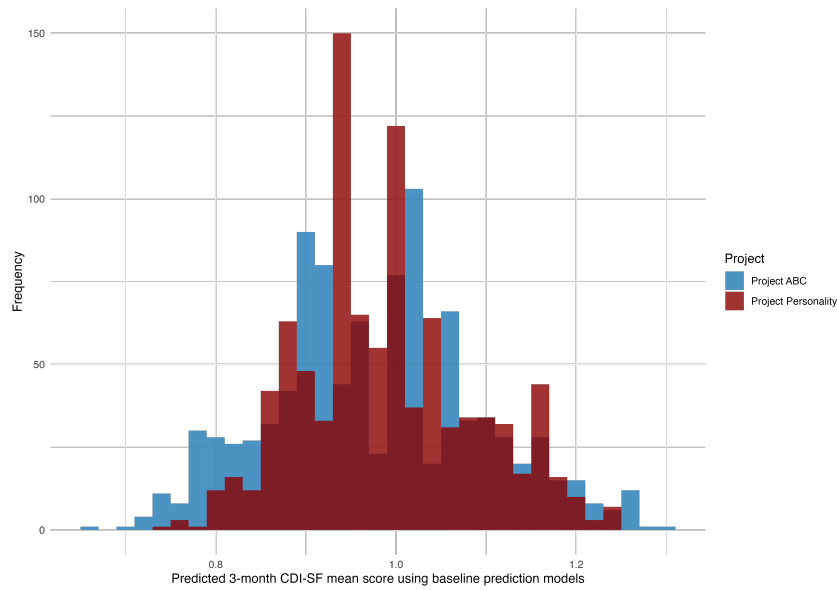
Table 2: Baseline depression severity prediction model results

Model[1]	Predictors	Estimates[2]	95% CI[3]	p-value
Project ABC[4]	Race [Black/African-American]	-0.04	[-0.16, 0.08]	0.527
	Race [Hispanic/Latinx]	-0.08	[-0.19, 0.02]	0.129
	Race [Mixed]	-0.03	[-0.13, 0.08]	0.617
	Race [White]	0.03	[-0.05, 0.12]	0.462
	Biological sex [Male]	-0.15	[-0.25, -0.04]	**0.007**
	Sexual orientation [LGBTQ+]	0.09	[0.02, 0.16]	**0.010**
	Sexual orientation [Other]	0.12	[0.03, 0.22]	**0.008**
	Gender identity [Women/girls]	-0.13	[-0.20, -0.06]	**<0.001**
	Gender identity [Male/Masculine]	-0.02	[-0.12, 0.08]	0.687
	Family challenges [Other]	0.07	[-0.01, 0.15]	0.070
	Family challenges [School]	0.03	[-0.03, 0.10]	0.313
	Coping strategies [No action]	-0.06	[-0.14, 0.02]	0.137
	Coping strategies [Positive]	-0.16	[-0.24, -0.08]	**<0.001**
Project Personality[5]	Race [Black/African-American]	-0.06	[-0.18, 0.07]	0.387
	Race [Hispanic/Latinx]	-0.05	[-0.16, 0.06]	0.382
	Race [Mixed]	-0.03	[-0.14, 0.08]	0.552
	Race [White]	-0.00	[-0.09, 0.09]	0.937
	Biological sex [Male]	-0.11	[-0.22, 0.01]	0.066
	Sexual orientation [LGBTQ+]	0.07	[0.00, 0.14]	**0.046**
	Sexual orientation [Other]	0.06	[-0.03, 0.16]	0.182
	Gender identity [Women/girls]	-0.15	[-0.22, -0.08]	**<0.001**
	Gender identity [Male/Masculine]	-0.10	[-0.21, -0.00]	**0.049**
	Family challenges [Other]	0.02	[-0.06, 0.10]	0.625
	Family challenges [School]	-0.02	[-0.09, 0.05]	0.664
	Coping strategies [No action]	-0.06	[-0.15, 0.03]	0.194
	Coping strategies [Positive]	-0.13	[-0.22, -0.04]	**0.005**

[1] Unadjusted for treatment assignment, and no statistical interactions between model variables were specified in the baseline depression severity prediction model; [2] Mean difference of each variable holding other variables constant; [3] Confidence interval; [4] Baseline prediction model constructed based on participants in Project ABC and Placebo Control groups; [5] Baseline prediction model constructed based on participants in Project Personality and Placebo Control groups.

The demographic predictors were used to estimate the baseline severity of depression for each individual in these two datasets using the constructed baseline prediction models. The distributions of the

predicted baseline depression severity scores for the two subsets are shown in Figure 2. These distributions indicate variability in the severity of depression among individuals before treatment assignment, with slightly greater variation observed in the Project ABC samples compared to the Project Personality samples. The predicted 3-month CDI-SF mean scores were ranked and divided into 10 baseline severity groups using deciles. The demographic characteristics were then compared across 10 levels of predicted CDI-SF mean score groups. The results showed that Non-binary individuals, LGBTQ+ individuals and participants who did not use any coping strategies were more commonly found in the high depression score groups, indicating individual's gender identity, sexual orientation and coping strategy contribute to the difference in depression severity. Detailed results can be found in Appendix Table 6 and Appendix Table 7.

Figure 2: The distributions of the predicted 3-month CDI-SF mean scores

The model performance is summarised in Table 3. For the Project ABC baseline depression severity prediction model, the model performed similarly in both the validation and test sets, with an RMSE of approximately 0.4, a calibration slope close to 1, and calibration-in-the-large close to 0. For the Project Personality baseline prediction model, the external validation results indicated that the calibration slope was not close to 1 and calibration-in-the-large was not close to 0, suggesting that the predictive ability of the model's predictors is limited.

Table 3: Performance metrics of baseline prediction models

| Model	Metric	Train-Validation[1]	Test[2]
	RMSE[3]	0.399	0.396
Project ABC	Calibration slope	1.158	0.923
	Calibration at large	-0.154	0.067
	RMSE	0.406	0.423
Project Personality	Calibration slope	1.200	0.056
	Calibration at large	-0.198	0.935

[1] Performance calculated from the derivation cohort using leave-one-out cross-validation framework; [2] Performance calculated based on the test dataset; [3] root mean squared error

3.3 Heterogeneity of Treatment Effect

The effect of each SSI on depression for each predicted depression mean score decile is shown in Figure 3. Negative mean differences indicate that the intervention is favored over the Placebo Control. The upper panel shows the results for Project ABC treatment effect. The ATE of Project ABC is -0.080 (s.d.: 0.022; p-value: <0.001) over samples in Project ABC and control groups. The interaction term of the HTE model is significant (coef: 0.393, p value: 0.039). The mean difference in the lowest score group is -0.16 (bootstrapped CI: [-0.20-0.04]), indicating a significant reduction in depression for those with the lowest baseline severity in Project ABC compared to the control group. Score group 9 and 10, show a mean difference of -0.03 (bootstrapped CI:[-0.13,0.02]) and 0.00 (bootstrapped CI: -0.14, 0.06) respectively, indicating no significant effect for those with the two most highest baseline depression severity. The treatment effects in score groups 6 through 8 were smaller than the ATE, whereas score groups 5 and below showed greater benefits than the ATE. Additionally, as the baseline severity increases, the mean differences generally become less negative, suggesting a diminishing treatment effect. These findings indicates that there is clear heterogeneity in the treatment effect of Project ABC.

The lower panel of Figure 3 shows the HTE for Project Personality. The ATE is -0.069 (s.d.: 0.023; p-value: <0.001). The interaction term is not statistically significant (p-value: 0.374). However, the stratification results show that the treatment effect varies across different score groups. Similar to Project ABC, the mean differences become less negative as baseline depression severity increases. Specifically, the ATE reflects the treatment effect among score groups 5 and 6. Participants in score groups 9 and 10 do not benefit from Project Personality, with mean differences of -0.04 (bootstrapped CI: [-0.16, 0.02]) and -0.03 (bootstrapped CI: [-0.18, 0.04]), respectively.
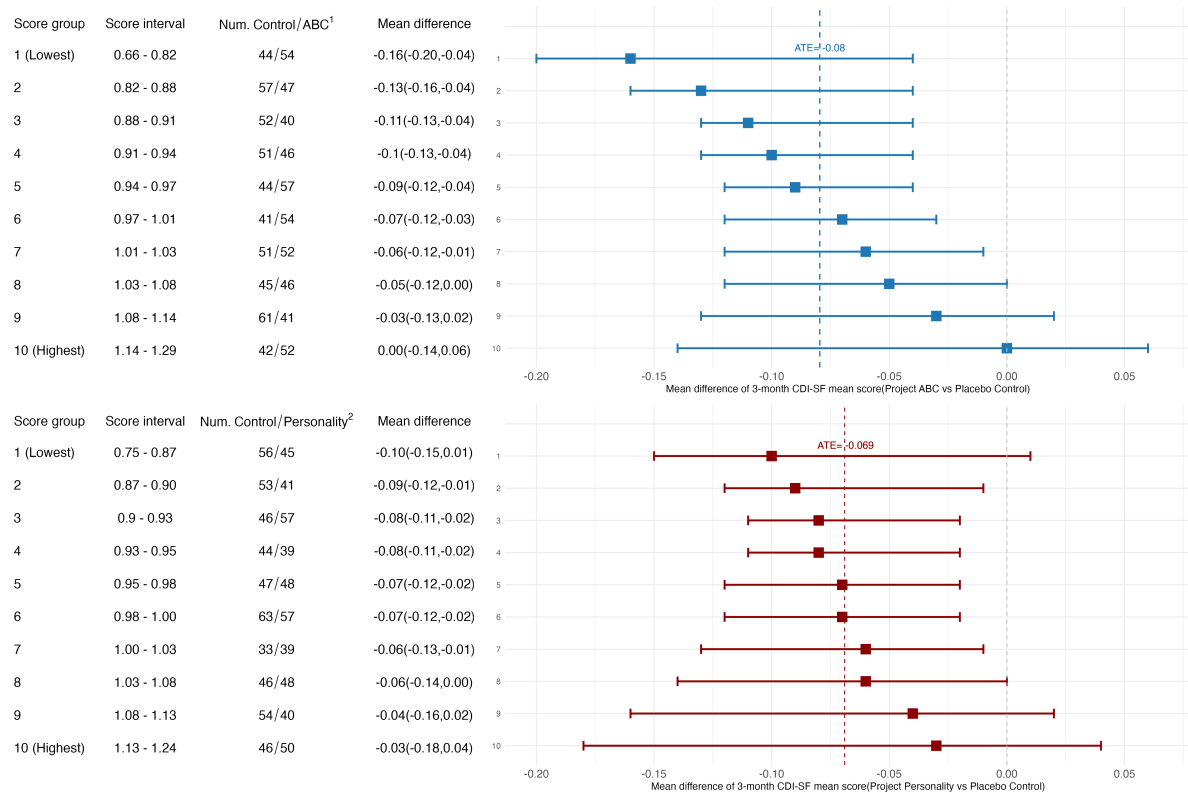
Figure 3: Heterogeneity of treatment effect using risk-based approach. The upper panel is for Project ABC, lower panel for Project Personality. [1] number of subjects in Placebo Control/number of subjects in Project ABC within each decile; [2] number of subjects in Placebo Control/number of subjects in Project Personality within each decile.

To summarise, both online SSIs show stronger effects in participants with lower baseline severity of depression. Though the HTE for Project Personality was not statistically significant. The treatment effect decreases as the baseline severity increases, with the highest baseline severity groups showing little to no significant treatment effect. This indicates that both interventions are more effective for individuals with milder forms of depression, while those with more severe baseline symptoms might require additional or alternative treatments for significant improvement.

4   Discussion

This study evaluated the heterogeneity of treatment effects for two online SSIs, Project ABC and Project Personality, in reducing depressive symptoms among adolescents. The baseline depression severity prediction models were constructed using demographic predictors, and the model performance was assessed. The predicted results were used as a potential effect modifier. This study identified variability in the treatment effects based on the baseline severity of depression.

This study has two key findings. First, the baseline depression severity prediction model suggested that subgroups of individuals whose gender identity is non-binary, who self-identified as LGBTQ+, and who did not use any coping strategies are more likely to develop severe depression and benefit less from the interventions. Second, the HTE for Project ABC is statistically significant, whereas it is not significant for Project Personality. However, risk stratification visualization suggested that HTE exists, and the effect sizes across different score groups differ from the ATE. Both interventions exhibited a stronger treatment effect in participants with lower baseline severity, and the treatment effect decreased with increasing baseline severity, showing no significant effect in the highest score group. Therefore, individuals with mild depression would benefit more from these two interventions.

This study has several strengths. This study used a risk-based approach to evaluate HTE. This approach has several advantages over conventional subgrouop analysis. First, the baseline depression severity score was included in the final HTE model as a continuous variable and preserved the variable information, while subgroup analysis estimates average treatment effect within a specific group. Second, the models were constructed using the entire dataset and have more statistical power in comparison to subgroup analysis. Third, the risk-based approach allows the estimation of the absolute value of depression score for each individuals, which help guide clinical decision-making using individual "risk" profiles. Lastly, the baseline prediction models were built under a LOOCV framework, which ensured the robustness of the prediction models.

This study also has a few limitations. First, the gender identity and coping strategies variables were re-coded using latent classes identified from LCA, where each individual was assigned to the latent class for which they had the highest probability. Some researchers have pointed out that this approach might be less effective in detecting heterogeneous treatment effects compared to more complex model-based approaches due to potential misclassification rates [8]. Second, the PATH (Predictive Approaches to Treatment Effect Heterogeneity) statement [12, 17] recommends using a high-quality, externally-developed, compatible model to identify subgroups when available. This study used an internally developed baseline prediction model based on available clinical characteristics. Therefore, the validity of the stratification results depends heavily on the validity of the baseline prediction model. Specifically, the external validation results for the Project Personality baseline prediction model indicated that the calibration slope and calibration-in-the-large were not ideal, suggesting limited predictive ability of the model's predictors. This may affect the accuracy of the stratification and, consequently, the interpretation of the HTE results for Project Personality. This may explain why the statistical interaction term for the Project Personality model was not significant. Additionally, the model assumed a linear relationship among predictors

and the outcome, which may not fully capture the complexities of depression severity. Moreover, the variables considered in the model are self-reported, which introduces a potential bias due to inaccuracies or inconsistencies in self-reporting. Lastly, there may exist unmeasured clinical characteristics such as genetic predispositions that may contribute to depression but were not captured in the current model. This may affect the accuracy of the predictions and the understanding of adolescent depression.

In conclusion, this study suggests there is heterogeneity in the treatment effects of both SSIs for adolescent depression. The use of a risk-based approach enables the estimation of individual depression scores, providing quantitative evidence for clinical practice and supporting personalized treatment assignment.

References

[1] Jessica L. Schleider and John R. Weisz. "Little Treatments, Promising Effects? Meta-Analysis of Single-Session Interventions for Youth Psychiatric Problems". In: *Journal of the American Academy of Child amp; Adolescent Psychiatry* 56.2 (Feb. 2017), pp. 107–115. ISSN: 0890-8567. DOI: 10.1016/j.jaac.2016.11.007. URL: http://dx.doi.org/10.1016/j.jaac.2016.11.007.

[2] Jessica L. Schleider et al. "Randomized Trial of a Single-Session Growth Mind-Set Intervention for Rural Adolescents' Internalizing and Externalizing Problems". In: *Journal of Clinical Child amp; Adolescent Psychology* 49.5 (June 2019), pp. 660–672. ISSN: 1537-4424. DOI: 10.1080/15374416.2019.1622123. URL: http://dx.doi.org/10.1080/15374416.2019.1622123.

[3] Jessica L. Schleider et al. "A randomized trial of online single-session interventions for adolescent depression during COVID-19". In: *Nature Human Behaviour* 6.2 (Dec. 2021), pp. 258–268. ISSN: 2397-3374. DOI: 10.1038/s41562-021-01235-0. URL: http://dx.doi.org/10.1038/s41562-021-01235-0.

[4] Guido W. Imbens and Donald B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction.* Cambridge University Press, Apr. 2015. ISBN: 9781139025751. DOI: 10.1017/cbo9781139025751. URL: http://dx.doi.org/10.1017/CBO9781139025751.

[5] Theodore J. Iwashyna et al. "Implications of Heterogeneity of Treatment Effect for Reporting and Analysis of Randomized Trials in Critical Care". In: *American Journal of Respiratory and Critical Care Medicine* 192.9 (Nov. 2015), pp. 1045–1051. ISSN: 1535-4970. DOI: 10.1164/rccm.201411-2125cp. URL: http://dx.doi.org/10.1164/rccm.201411-2125CP.

[6] Kurt Kroenke, Robert L. Spitzer, and Janet B. W. Williams. "The Patient Health Questionnaire-2: Validity of a Two-Item Depression Screener". In: *Medical Care* 41.11 (Nov. 2003), pp. 1284–1292. ISSN: 0025-7079. DOI: 10.1097/01.mlr.0000093487.78664.3c. URL: http://dx.doi.org/10.1097/01.MLR.0000093487.78664.3C.

[7] Eric Stice et al. "Meta-analytic review of dissonance-based eating disorder prevention programs: Intervention, participant, and facilitator features that predict larger effects". In: *Clinical Psychology Review* 70 (June 2019), pp. 91–107. ISSN: 0272-7358. DOI: 10.1016/j.cpr.2019.04.004. URL: http://dx.doi.org/10.1016/j.cpr.2019.04.004.

[8] Stephanie T. Lanza and Brittany L. Rhoades. "Latent Class Analysis: An Alternative Perspective on Subgroup Analysis in Prevention and Treatment". In: *Prevention Science* 14.2 (Feb. 2011), pp. 157–168. ISSN: 1573-6695. DOI: 10.1007/s11121-011-0201-1. URL: http://dx.doi.org/10.1007/s11121-011-0201-1.

[9] Pratik Sinha, Carolyn S. Calfee, and Kevin L. Delucchi. "Practitioner's Guide to Latent Class Analysis: Methodological Considerations and Common Pitfalls". In: *Critical Care Medicine* 49.1 (Nov. 2020), e63–e79. ISSN: 0090-3493. DOI: 10.1097/ccm.0000000000004710. URL: http://dx.doi.org/10.1097/CCM.0000000000004710.

[10] Drew A. Linzer and Jeffrey B. Lewis. "poLCA: AnRPackage for Polytomous Variable Latent Class Analysis". In: *Journal of Statistical Software* 42.10 (2011). ISSN: 1548-7660. DOI: 10.18637/jss.v042.i10. URL: http://dx.doi.org/10.18637/jss.v042.i10.

[11] David M. Kent and Rodney A. Hayward. "Limitations of Applying Summary Results of Clinical Trials to Individual Patients: The Need for Risk Stratification". In: *JAMA* 298.10 (Sept. 2007), p. 1209. ISSN: 0098-7484. DOI: 10.1001/jama.298.10.1209. URL: http://dx.doi.org/10.1001/jama.298.10.1209.

[12] David M. Kent et al. "The Predictive Approaches to Treatment effect Heterogeneity (PATH) Statement". In: *Annals of Internal Medicine* 172.1 (Nov. 2019), p. 35. ISSN: 0003-4819. DOI: 10.7326/m18-3667. URL: http://dx.doi.org/10.7326/M18-3667.

[13] Max Kuhn. *caret: Classification and Regression Training*. R package version 6.0-94. 2023. URL: https://CRAN.R-project.org/package=caret.

[14] Ardalan Mirzaei et al. "Missing data in surveys: Key concepts, approaches, and applications". In: *Research in Social and Administrative Pharmacy* 18.2 (Feb. 2022), pp. 2308–2316. ISSN: 1551-7411. DOI: 10.1016/j.sapharm.2021.03.009. URL: http://dx.doi.org/10.1016/j.sapharm.2021.03.009.

[15] Shaun R Seaman and Ian R White. "Review of inverse probability weighting for dealing with missing data". In: *Statistical Methods in Medical Research* 22.3 (Jan. 2011), pp. 278–295. ISSN: 1477-0334. DOI: 10.1177/0962280210395740. URL: http://dx.doi.org/10.1177/0962280210395740.

[16] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2024. URL: https://www.R-project.org/.

[17] David M. Kent et al. "The Predictive Approaches to Treatment effect Heterogeneity (PATH) Statement: Explanation and Elaboration". In: *Annals of Internal Medicine* 172.1 (Nov. 2019), W1. ISSN: 0003-4819. DOI: 10.7326/m18-3668. URL: http://dx.doi.org/10.7326/M18-3668.

5 Appendix

5.1 Sensitivity Analysis

To test the robustness of the study results, a sensitivity analysis was conducted using imputed data. First, the extent of missing data was investigated. Three variables had missing data: sexual orientation (6, 0.4%), biological sex (22, 1.47%), and race (33, 2.2%). Figures 4 and 5 display the missing data patterns by observations and by variables, respectively. The results indicate no specific pattern of missingness.
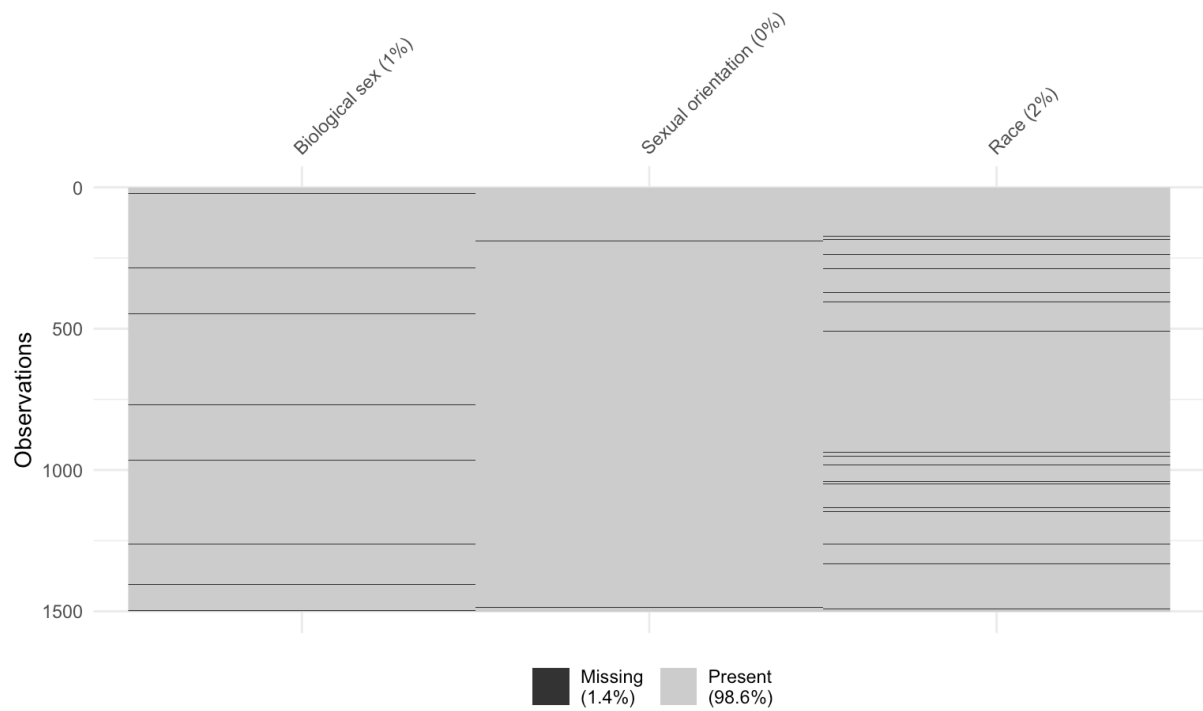
Figure 4: Missing pattern by observations
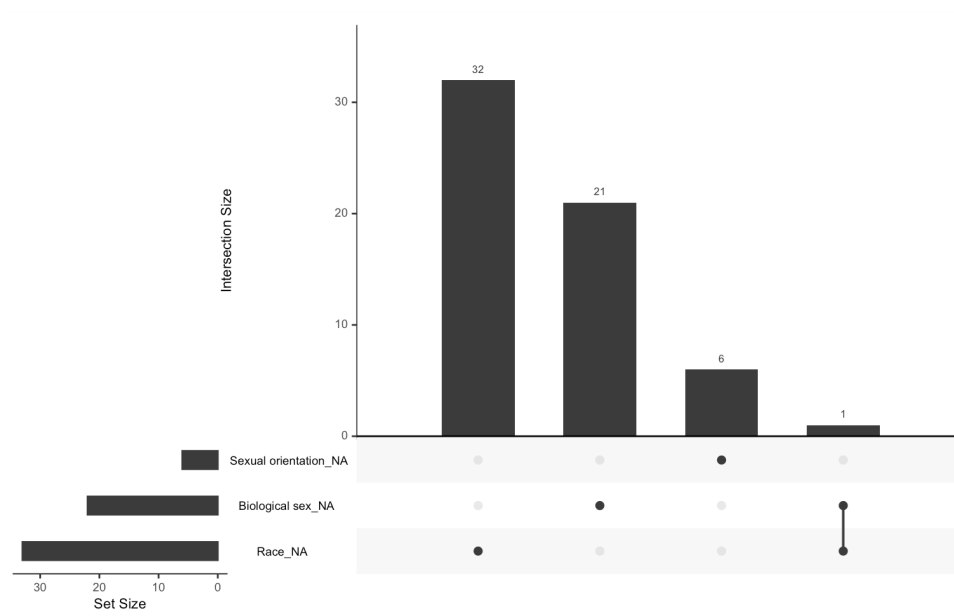
Figure 5: Missing pattern by variables

Inverse probability missing weights were applied to account for missing. The risk stratification results, summarised in Figure 6, were similar to the CCA results.
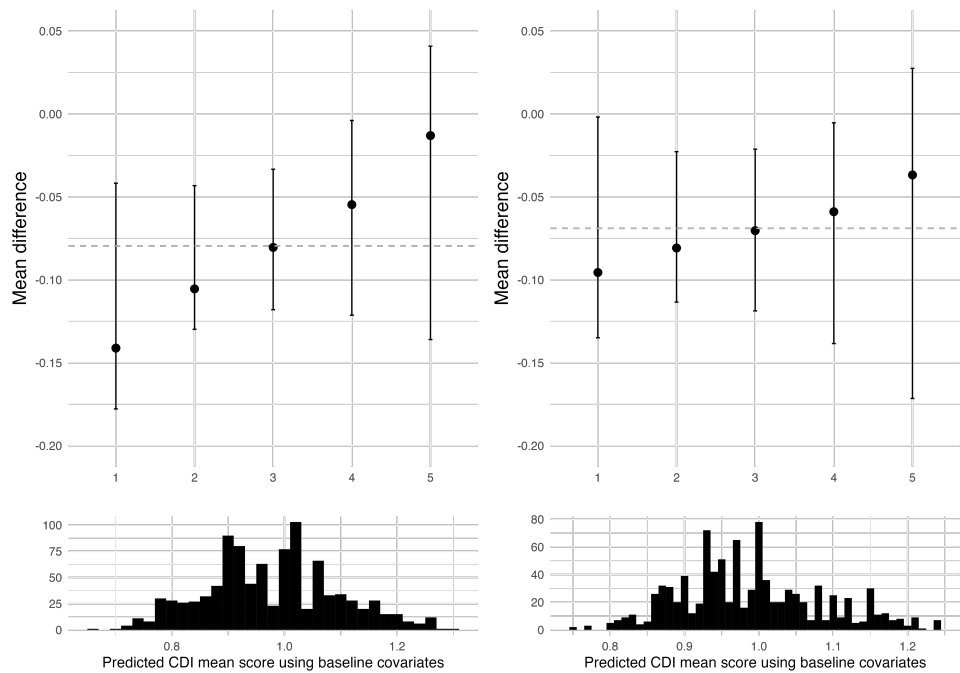
(A) CCA stratification results
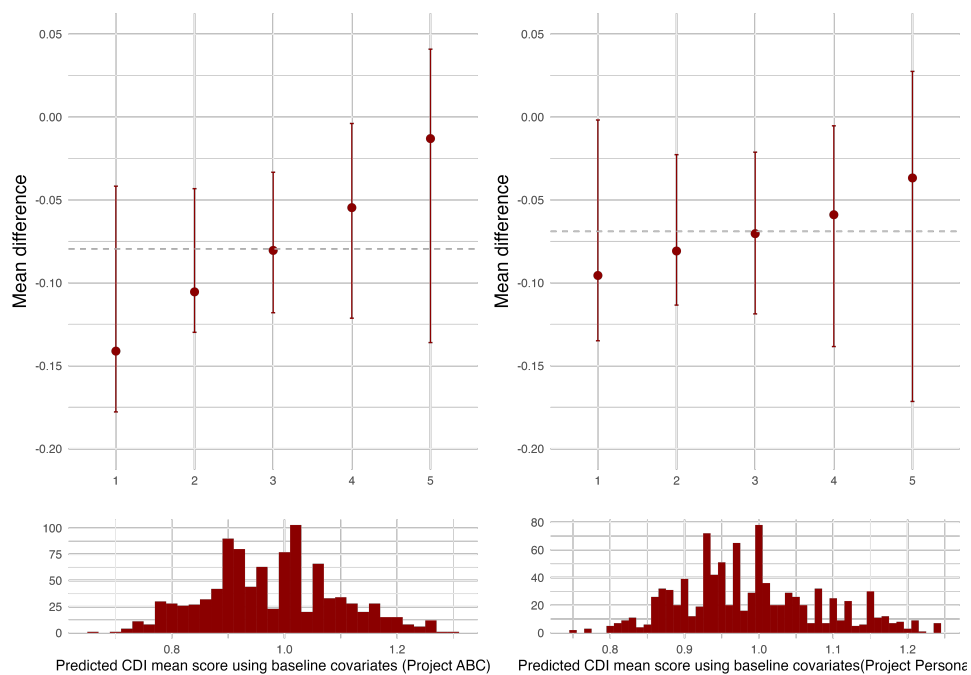
(B) IPCW-based stratification results

Figure 6: Overall caption for the figure

5.2 Appendix Tables

Table 4: Variable dictionary

Variable	Possible Values
condition	treatment arm allocation
baseline CDI-SF score	baseline CDI-SF score
f1 CDI-SF score	3-month follow-up CDI-SF score (**primary outcome**)
Biological sex	Male
	Female
	Other
Age	Numeric (years)
Race/Ethnicity	American Indian or Alaska Native
	Asian including Asian Desi
	Hispanic/Latinx
	Native Hawaiian or Other Pacific Islander
	White/Caucasian non-Hispanic (includes Middle Eastern)
	Black/African American
	(can select multiple)
Sexual orientation	Pansexual
	Other/Not listed (please specify)
	Gay/Lesbian/Homosexual
	Bisexual
	Heterosexual/Straight
	Asexual
	Queer
	I do not use a label
	Unsure/Questioning
	I do not want to respond
Family challenges during COVID-19 pandemic	Not enough money for food
	No regular place to sleep or stay
	Could not attend school in person
	Could not attend school at all
	Other
	(can select multiple)
Coping strategies during COVID-19 pandemic	Connecting with others
	Talking with people you trust
	Contacting a healthcare provider
	Drinking alcohol
	Smoking more cigarettes or vaping more
	(can select multiple)

Table 4: Variable dictionary (continued)

Variable	Possible Values
Gender identity	Agender
	Not sure
	Other (please specify)
	Androgynous
	Nonbinary
	Two-Spirited
	Female to Male Transgender (FTM)
	Trans Female/Trans Feminine
	Trans Male/Trans Masculine
	Gender Expansive
	Third Gender
	Genderqueer
	Male to Female Transgender (MTF)
	Man/Boy
	Transgender
	Woman/Girl
	(can select multiple)

Table 5: Likelihood ratio test of each variables in the baseline prediction models

	Project ABC			Project Personality		
	Deviance	DF	P-value	Deviance	DF	P-value
Gender identity	15.456260	2	< 0.001	16.358782	2	< 0.001
Race	9.522943	4	0.05	2.266644	4	0.69
Biological sex	7.509756	1	0.01	3.441522	1	0.06
Sexual orientation	8.728432	2	0.01	4.084130	2	0.13
Challenges	3.339847	2	0.19	1.220614	2	0.54
Coping strategies	19.418374	2	< 0.001	10.608348	2	< 0.001

Table 6: Demographic Characteristics stratified by predicted 3-month CDI-SF mean score groups among patients received Project ABC and Placebo Control treatment

Demographics	1 N = 102	2 N = 107	3 N = 96	4 N = 123	5 N = 98	6 N = 83	7 N = 98	8 N = 75	9 N = 97	10 N = 98
condition										
Placebo Control	42 (41%)	52 (49%)	51 (53%)	57 (46%)	53 (54%)	42 (51%)	51 (52%)	44 (59%)	53 (55%)	43 (44%)
Project ABC	60 (59%)	55 (51%)	45 (47%)	66 (54%)	45 (46%)	41 (49%)	47 (48%)	31 (41%)	44 (45%)	55 (56%)
Baseline CDI mean score(0-2)	1.20 (0.34)	1.09 (0.35)	1.19 (0.33)	1.17 (0.35)	1.19 (0.35)	1.15 (0.33)	1.10 (0.37)	1.12 (0.35)	1.16 (0.30)	1.17 (0.35)
Race										
Asian Including Asian Desi	13 (13%)	16 (15%)	8 (8.3%)	17 (14%)	13 (13%)	11 (13%)	10 (10%)	5 (6.7%)	7 (7.2%)	8 (8.2%)
Black/African-American	7 (6.9%)	7 (6.5%)	10 (10%)	7 (5.7%)	7 (7.1%)	2 (2.4%)	5 (5.1%)	8 (11%)	11 (11%)	9 (9.2%)
Hispanic/Latinx	13 (13%)	11 (10%)	9 (9.4%)	14 (11%)	12 (12%)	11 (13%)	19 (19%)	11 (15%)	9 (9.3%)	9 (9.2%)
Mixed	14 (14%)	10 (9.3%)	12 (13%)	23 (19%)	8 (8.2%)	11 (13%)	22 (22%)	11 (15%)	21 (22%)	10 (10%)
White	55 (54%)	63 (59%)	57 (59%)	62 (50%)	58 (59%)	48 (58%)	42 (43%)	40 (53%)	49 (51%)	62 (63%)
Age (yrs)										
13	10 (9.8%)	4 (3.7%)	8 (8.3%)	7 (5.7%)	6 (6.1%)	7 (8.4%)	4 (4.1%)	3 (4.0%)	7 (7.2%)	4 (4.1%)
14	14 (14%)	23 (21%)	12 (13%)	16 (13%)	15 (15%)	14 (17%)	21 (21%)	10 (13%)	18 (19%)	15 (15%)
15	35 (34%)	26 (24%)	38 (40%)	43 (35%)	37 (38%)	20 (24%)	22 (22%)	27 (36%)	25 (26%)	33 (34%)
16	43 (42%)	54 (50%)	38 (40%)	57 (46%)	40 (41%)	42 (51%)	51 (52%)	35 (47%)	47 (48%)	46 (47%)
Biological sex										
Female	91 (89%)	92 (86%)	86 (90%)	109 (89%)	87 (89%)	77 (93%)	88 (90%)	67 (89%)	85 (88%)	89 (91%)
Male	11 (11%)	15 (14%)	10 (10%)	14 (11%)	11 (11%)	6 (7.2%)	10 (10%)	8 (11%)	12 (12%)	9 (9.2%)
Sexual orientation										
Heterosexual	23 (23%)	23 (21%)	16 (17%)	24 (20%)	23 (23%)	15 (18%)	27 (28%)	13 (17%)	18 (19%)	23 (23%)
LGBTQ	65 (64%)	68 (64%)	68 (71%)	81 (66%)	63 (64%)	55 (66%)	63 (64%)	52 (69%)	67 (69%)	54 (55%)
Other	14 (14%)	16 (15%)	12 (13%)	18 (15%)	12 (12%)	13 (16%)	8 (8.2%)	10 (13%)	12 (12%)	21 (21%)
Language										
English	97 (95%)	105 (98%)	96 (100%)	121 (98%)	95 (97%)	79 (95%)	93 (95%)	74 (99%)	96 (99%)	96 (98%)
Other	5 (4.9%)	2 (1.9%)	0 (0%)	2 (1.6%)	3 (3.1%)	4 (4.8%)	5 (5.1%)	1 (1.3%)	1 (1.0%)	2 (2.0%)
Gender identity										
Non-binary	20 (20%)	26 (24%)	21 (22%)	22 (18%)	25 (26%)	16 (19%)	16 (16%)	15 (20%)	20 (21%)	19 (19%)
Women/girls	71 (70%)	63 (59%)	62 (65%)	82 (67%)	59 (60%)	58 (70%)	69 (70%)	49 (65%)	59 (61%)	66 (67%)
Male/Masculine	11 (11%)	18 (17%)	13 (14%)	19 (15%)	14 (14%)	9 (11%)	13 (13%)	11 (15%)	18 (19%)	13 (13%)
Number of challenges										
0	18 (18%)	22 (21%)	15 (16%)	22 (18%)	22 (22%)	13 (16%)	27 (28%)	13 (17%)	25 (26%)	20 (20%)
1	62 (61%)	63 (59%)	59 (61%)	78 (63%)	55 (56%)	57 (69%)	49 (50%)	42 (56%)	50 (52%)	58 (59%)
>=2	22 (22%)	22 (21%)	22 (23%)	23 (19%)	21 (21%)	13 (16%)	22 (22%)	20 (27%)	22 (23%)	20 (20%)

Type of challenges										
No impact	18 (18%)	22 (21%)	15 (16%)	22 (18%)	22 (22%)	13 (16%)	27 (28%)	13 (17%)	25 (26%)	20 (20%)
Other	27 (26%)	23 (21%)	23 (24%)	26 (21%)	18 (18%)	14 (17%)	23 (23%)	20 (27%)	20 (21%)	22 (22%)
School	57 (56%)	62 (58%)	58 (60%)	75 (61%)	58 (59%)	56 (67%)	48 (49%)	42 (56%)	52 (54%)	56 (57%)
Type of coping strategies										
Combined	10 (9.8%)	11 (10%)	7 (7.3%)	24 (20%)	7 (7.1%)	13 (16%)	14 (14%)	13 (17%)	17 (18%)	3 (3.1%)
No action	48 (47%)	51 (48%)	48 (50%)	60 (49%)	50 (51%)	38 (46%)	47 (48%)	35 (47%)	37 (38%)	49 (50%)
Positive	44 (43%)	45 (42%)	41 (43%)	39 (32%)	41 (42%)	32 (39%)	37 (38%)	27 (36%)	43 (44%)	46 (47%)

Table 7: Demographic Characteristics stratified by predicted 3-month CDI-SF mean score groups among patients received Project Personality and Placebo Control treatment

Demographics	1	2	3	4	5	6	7	8	9	10
	N = 101	N = 94	N = 103	N = 83	N = 95	N = 120	N = 72	N = 94	N = 94	N = 96
condition										
Placebo Control	56 (55%)	53 (56%)	46 (45%)	44 (53%)	47 (49%)	63 (53%)	33 (46%)	46 (49%)	54 (57%)	46 (48%)
Project Personality	45 (45%)	41 (44%)	57 (55%)	39 (47%)	48 (51%)	57 (48%)	39 (54%)	48 (51%)	40 (43%)	50 (52%)
Baseline CDI mean score(0-2)	0.93 (0.32)	1.02 (0.39)	1.06 (0.33)	1.12 (0.37)	1.19 (0.31)	1.19 (0.28)	1.25 (0.30)	1.24 (0.36)	1.29 (0.33)	1.35 (0.30)
Race										
Asian Including Asian Desi	11 (11%)	9 (9.6%)	3 (2.9%)	26 (31%)	15 (16%)	5 (4.2%)	15 (21%)	2 (2.1%)	8 (8.5%)	6 (6.3%)
Black/African-American	14 (14%)	12 (13%)	12 (12%)	8 (9.6%)	5 (5.3%)	4 (3.3%)	3 (4.2%)	2 (2.1%)	6 (6.4%)	3 (3.1%)
Hispanic/Latinx	20 (20%)	26 (28%)	8 (7.8%)	0 (0%)	24 (25%)	13 (11%)	5 (6.9%)	3 (3.2%)	8 (8.5%)	3 (3.1%)
Mixed	9 (8.9%)	23 (24%)	10 (9.7%)	9 (11%)	12 (13%)	9 (7.5%)	19 (26%)	17 (18%)	15 (16%)	14 (15%)
White	47 (47%)	24 (26%)	70 (68%)	40 (48%)	39 (41%)	89 (74%)	30 (42%)	70 (74%)	57 (61%)	70 (73%)
Age (yrs)										
13	3 (3.0%)	4 (4.3%)	1 (1.0%)	3 (3.6%)	10 (11%)	9 (7.5%)	7 (9.7%)	7 (7.4%)	5 (5.3%)	7 (7.3%)
14	15 (15%)	11 (12%)	18 (17%)	15 (18%)	9 (9.5%)	15 (13%)	13 (18%)	13 (14%)	14 (15%)	17 (18%)
15	39 (39%)	28 (30%)	34 (33%)	25 (30%)	31 (33%)	37 (31%)	23 (32%)	32 (34%)	35 (37%)	28 (29%)
16	44 (44%)	51 (54%)	50 (49%)	40 (48%)	45 (47%)	59 (49%)	29 (40%)	42 (45%)	40 (43%)	44 (46%)
Biological sex										
Female	58 (57%)	77 (82%)	94 (91%)	78 (94%)	87 (92%)	113 (94%)	71 (99%)	87 (93%)	91 (97%)	96 (100%)
Male	43 (43%)	17 (18%)	9 (8.7%)	5 (6.0%)	8 (8.4%)	7 (5.8%)	1 (1.4%)	7 (7.4%)	3 (3.2%)	0 (0%)
Sexual orientation										
Heterosexual	86 (85%)	44 (47%)	29 (28%)	27 (33%)	18 (19%)	7 (5.8%)	3 (4.2%)	0 (0%)	0 (0%)	0 (0%)
LGBTQ	13 (13%)	45 (48%)	66 (64%)	51 (61%)	63 (66%)	87 (73%)	58 (81%)	76 (81%)	75 (80%)	66 (69%)
Other	2 (2.0%)	5 (5.3%)	8 (7.8%)	5 (6.0%)	14 (15%)	26 (22%)	11 (15%)	18 (19%)	19 (20%)	30 (31%)
Language										
English	94 (93%)	90 (96%)	100 (97%)	80 (96%)	92 (97%)	119 (99%)	70 (97%)	94 (100%)	93 (99%)	94 (98%)
Other	7 (6.9%)	4 (4.3%)	3 (2.9%)	3 (3.6%)	3 (3.2%)	1 (0.8%)	2 (2.8%)	0 (0%)	1 (1.1%)	2 (2.1%)
Gender identity										
Non-binary	0 (0%)	0 (0%)	2 (1.9%)	0 (0%)	2 (2.1%)	3 (2.5%)	3 (4.2%)	20 (21%)	67 (71%)	91 (95%)
Women/girls	61 (60%)	77 (82%)	93 (90%)	78 (94%)	82 (86%)	95 (79%)	61 (85%)	55 (59%)	17 (18%)	0 (0%)
Male/Masculine	40 (40%)	17 (18%)	8 (7.8%)	5 (6.0%)	11 (12%)	22 (18%)	8 (11%)	19 (20%)	10 (11%)	5 (5.2%)
Number of challenges										

0	10 (9.9%)	27 (29%)	4 (3.9%)	37 (45%)	20 (21%)	15 (13%)	25 (35%)	11 (12%)	14 (15%)	20 (21%)
1	78 (77%)	56 (60%)	77 (75%)	37 (45%)	48 (51%)	85 (71%)	30 (42%)	58 (62%)	43 (46%)	39 (41%)
>=2	13 (13%)	11 (12%)	22 (21%)	9 (11%)	27 (28%)	20 (17%)	17 (24%)	25 (27%)	37 (39%)	37 (39%)
Type of challenges										
No impact	10 (9.9%)	27 (29%)	4 (3.9%)	37 (45%)	20 (21%)	15 (13%)	25 (35%)	11 (12%)	14 (15%)	20 (21%)
Other	11 (11%)	10 (11%)	21 (20%)	9 (11%)	37 (39%)	19 (16%)	19 (26%)	39 (41%)	41 (44%)	37 (39%)
School	80 (79%)	57 (61%)	78 (76%)	37 (45%)	38 (40%)	86 (72%)	28 (39%)	44 (47%)	39 (41%)	39 (41%)
Type of coping strategies										
Combined	0 (0%)	0 (0%)	0 (0%)	2 (2.4%)	1 (1.1%)	8 (6.7%)	15 (21%)	29 (31%)	25 (27%)	31 (32%)
No action	16 (16%)	28 (30%)	33 (32%)	36 (43%)	63 (66%)	95 (79%)	49 (68%)	50 (53%)	29 (31%)	65 (68%)
Positive	85 (84%)	66 (70%)	70 (68%)	45 (54%)	31 (33%)	17 (14%)	8 (11%)	15 (16%)	40 (43%)	0 (0%)