

An Investigation of the Impact of Neuroticism on the Efficacy of Mobile App-Based Interventions for
Generalized Anxiety Disorder

March 11, 2022

Jenna Bedrava

Abstract

The increasing prevalence of mental health disorders can be considered an epidemic, made worse by the mental healthcare system's inability to meet the demand for care. The University of Michigan's Providing Mental Health Precision Treatment (PROMPT) study enrolled 4,500 patients to a randomized trial of mobile health interventions for mental illness from the waitlist of patients waiting for initial intake appointments for behavioral health care at Michigan Medicine. Previous studies have found an association between one of the Big Five personality traits, neuroticism, and psychological disorders. In this investigation, we use generalized linear mixed effects models to determine if neuroticism score, as measured using the NEO-FFI personality test, is predictive of mobile health treatment efficacy with the goal of making recommendations for more efficacious, individualized treatments. We find that neuroticism is associated with GAD-7 score over time, as well as with the effect of the interventions as a whole. However, we do not find significant differences in each of the treatment effects or in neuroticism's impact on these individual effects. Due to the potential of confounding bias, as well as the presence of selection and self-reporting bias, we do not recommend these results be generalized to populations outside the study sample.

Introduction

The Global Burden of Disease Study estimated 45.82 million incident cases of anxiety disorders in 2019, a 50% increase from 1990 (Yang, Xiaorong, et al., 2019). With this increase in incidence comes an increased burden on society, as evidenced by an estimated 28.68 million disability-adjusted life years ascribable to anxiety disorders (Yang, Xiaorong, et al., 2019). The growing incidence of anxiety and other psychological disorders is worsened by healthcare systems' inability to meet the demand for treatment. It is estimated that 141 million, or 56% of, U.S. adults have sought or wanted to seek mental health

treatment for themselves or others, as of 2018. Of those who have tried to seek treatment, 38% of them report having to had wait longer than one week for services, yet in a survey about public opinions regarding mental health services, 81% of participants agree that patients should not have to wait longer than one week for care, as this can be dangerous to the individuals seeking help and those around them (2018).

Motivated by this exceedingly high prevalence of mental health disorders and our healthcare system's evident inability to meet the growing demand for care, investigators of the University of Michigan's Providing Mental Health Precision Treatment (PROMPT) study enrolled participants in a study of mobile health interventions for mental illness. The overarching goal of PROMPT is to reduce the burden of mental illness on healthcare systems by increasing the number of patients able to receive treatment at one time and by developing individualized pharmacological, psychological, and mobile-based treatments.

Moreover, neuroticism is defined as one's disposition to experiencing negative emotions, including anger, anxiety, self-consciousness, and emotional instability. The relationship between levels of neuroticism and anxiety disorder symptom severity has long been a topic of interest for mental illness researchers. It has been shown that neuroticism, as measured by the Neuroticism-Extraversion-Openness Five Factor Inventory, was the best predictor for all psychological problems, compared to the other 4 dimensions of personality (Rosellini & Brown, 2010). In this investigation, we aim to determine the potential impact of neuroticism score on the efficacy of the PROMPT study's mobile health app-based interventions for anxiety with the goal of assisting in the development of more efficacious, individualized mobile treatments that will lessen the current strain on the mental healthcare system and provide better care to patients in need.

Methods

Data Collection

PROMPT investigators recruited 4,500 participants from the waitlist of patients at Michigan Medicine waiting for intake appointments with a behavioral health clinician at least 2 weeks out from their scheduled appointment. Patients were randomly assigned to one of five treatment groups, all of which are combinations of different mobile app-based interventions. The three mobile applications used as treatments within this study include an app developed by PROMPT investigators called MyDataHelps, Headspace, and Silvercloud. Each participant was provided a FitBit, to which MyDataHelps connected in order to provide patients with measurements of mental and physical health markers, including step count, sleep quality, mood tracking, and mindfulness tracking. In this analysis, MyDataHelps is also referred to as Feedback, as this app provided daily notifications to encourage usage. Headspace is a personal meditation app designed to teach users mindful techniques. Similarly, Silvercloud is an app designed to teach users tactics to cope with behavioral health issues, including life balance, communication skills, and goal setting. Assigned treatment groups included: Feedback, Headspace, Silvercloud, Headspace plus Feedback, and Silvercloud plus Feedback.

Survey data from participants was collected at multiple time points and spans a wide range of variables, including demographic characteristics, NEO neuroticism score, Alcohol, Smoking, and Substance Involvement Screening Test (ASSIST) score, Generalized Anxiety Disorder (GAD-7) score, Interpersonal Evaluation List (ISEL) score, Positive and Negative Suicide Ideation (PANSI) score, Patient Health Questionnaire (PHQ) score, and more.

Outcome

To quantify anxiety symptom severity in study participants over time, we use the General Anxiety Disorder-7 (GAD-7) questionnaire, a screening tool based on DSM-V criteria for General Anxiety Disorder (GAD). The questionnaire consists of 8 questions, each of which are scored on a Likert scale ranging from 0 to 3. The first 7 items are summed for a total score, with higher scores representing more

intense GAD symptom severity. GAD-7 score can be categorized into 4 levels: minimal (GAD-7 scores between 0 and 4), mild (GAD-7 scores between 5 and 9), moderate (GAD-7 scores between 10 and 14), and severe (GAD-7 scores between 15 and 21). We calculate a GAD-7 score for each study participant at each of the 4 data collection timepoints and use these measurements as our primary outcome of interest in our analysis.

Covariate of Interest

The Neuroticism-Extraversion-Openness Five Factor Inventory (NEO-FFI) is a 60-item personality assessment that provides a measure of the Big Five dimensions of personality: neuroticism, extraversion, openness to experience, conscientiousness, and agreeableness. The assessment consists of 60 items answered on a 5-point Likert scale. Baseline survey data collected in the PROMPT study included the NEO-FFI items related to neuroticism. These items include statements such as “I often feel inferior to others,” “I feel comfortable in the presence of my bosses or other authorities,” and “I rarely feel anxious or nervous.” Items that were rated with a flipped Likert scale were converted and all items were summed for a total neuroticism score, ranging from 0 to 52. High values of neuroticism score indicate more neurotic personality types.

Statistical Methods

Summary statistics and visualizations are generated as part of exploratory data analysis. To elucidate the role neuroticism plays in the efficacy of mobile interventions on anxiety, we use generalized linear mixed models with a log link to model the change in GAD-7 score over time. We employ this link function to accommodate for the integer-only values of our outcome, GAD-7 score. We use random effects models to account for the correlation among repeated measures of the same individual. We adjust for demographic variables age, sex, race, and education and standardize all continuous variables, including neuroticism score and age, by subtracting the mean and dividing by the standard deviation. The Feedback only treatment group, male, white, and less than a high school

education are used as reference groups for treatment, sex, race, and education level, respectively.

Models with differing random effects and different fixed effects are compared using Likelihood Ratio Tests. A model with both a random intercept and a random slope for time is found to be a significantly better fit than a model with only a random intercept ($\chi^2 = 161.47$, p-value < 0.001). Model validity is then checked using diagnostic plots.

All analyses are performed in R Studio Version 1.4.1103. Generalized linear mixed models are fit using the lme4 package. A significance level of $\alpha = 0.05$ is employed for all analyses.

Results

Descriptive statistics of our sample, both overall and by quartile of neuroticism score, are provided in Table 1. We see that mean and median GAD-7 scores at all time points increase with each quartile of neuroticism score. In addition, mean and median GAD-7 scores decrease at each time of data collection. The majority of study population is female, white, and has higher than a high school education.

The distributions of our primary covariate and outcome of interest are visualized in Figure 1. The distribution of GAD-7 scores shifts downwards at each timepoint of data collection, implying curtailment of GAD symptoms throughout the duration of the study. We see that neuroticism score is approximately Normally distributed with a slight left skew. The mean neuroticism score is 34.48 with a standard deviation of 8.27. To visualize the relationship among GAD-7 and neuroticism, we use a side-by-side boxplot of neuroticism by each GAD-7 score category: minimal, mild, moderate and severe (Figure 2). We see that the distribution of neuroticism shifts upwards for each increase in GAD-7 score category, indicating that more intense GAD symptoms are associated with higher neuroticism levels. Interestingly, the range of neuroticism scores for each category does not span all possible values and also shifts upwards as GAD-7 category increases. Next, we visualize GAD-7 scores over time by treatment group

(Figure 3). All groups have similar trends in GAD-7 scores, but the group assigned to use only Headspace appears to be most efficacious in reducing anxiety symptoms.

Parameter estimates from our final generalized linear mixed effects model, fit with a log link, are given in Figure 4. This model equation is specified in Equation 1 in the Supplement. When performing model selection of fixed effects, we find that no interactions of treatment group and neuroticism were significant and using a Likelihood Ratio Test, a model not including these terms fit the data better ($\chi^2=4.38$, p-value = 0.357). Parameter estimates of the model including these interactions is provided in Supplement 1. From our final model, we find that the effect of neuroticism is significant ($e^{\beta}=1.35$, p-value <0.001), but the effects of all treatment groups are not significantly different than the Feedback only group (p-values all greater than 0.05). Because of ethical concerns, there is not a control group included in the PROMPT study. Therefore, inference cannot be made on the efficacy of the individual interventions compared to no intervention at all. The models fit in this analysis use the group of people who were only assigned to the PROMPT-based app, MyDataHelps, as the control group and thus, model parameters can only be interpreted as the difference in reduction of GAD-7 score compared to this control group. We find no significant differences.

All else being equal, we expect a 14% decrease in GAD-7 score in between each datapoint, implying that overall, the mobile health treatments are successful in reducing anxiety symptoms ($e^{\beta}=0.86$, p-value <0.001). Furthermore, there is an expected 35% increase in GAD-7 score for each additional neuroticism point, adjusting for all other covariates ($e^{\beta}=1.35$, p-value <0.001). Lastly, the estimated effect of neuroticism on the effect of time is not significant ($e^{\beta}=0.99$, p-value = 0.121), indicating that neuroticism does not impact treatment efficacy of the app-based treatments overall.

To assess homogeneity of variance, we examine a plot of residuals vs. fitted values for our final model (Figure 5). We see that our residuals are not randomly scattered and thus, this assumption cannot be assumed to be met. However, we do justify the trend depicted in residuals with our

assumption that GAD-7 score follows a Poisson distribution. We then assess normality of our model's random effects via Q-Q plots (Figure 6). We do not find any deviations from normality and assume this condition to be met.

Conclusion

All in all, using a generalized linear mixed effects model, it was found that neuroticism is significantly positively associated with GAD-7; more neurotic personalities are associated with higher GAD symptom severity. However, we did not find evidence of a significant relationship between neuroticism and treatment group. Hence, we cannot make any recommendations regarding individualized treatments for patients based on neuroticism score. However, there was significant evidence of a time effect, meaning that overall, the mobile health interventions employed in this study are effective in decreasing GAD symptom severity. We cannot make any conclusions about the advantage of one treatment over another.

It's important to discuss the utility and generalizability of these results. As mentioned earlier, patients are enrolled into the study from the waitlist of people waiting to see a behavioral health clinician. However, data about when and in what capacity clinical treatment began was not collected, and thus, this information is not included in our analysis. It is possible that the effect over time we are seeing is confounded by the delivery of clinical treatments. In addition, as mentioned previously, the assumption of homogeneity of variance was not quite met by our final model. In future investigations, we hope to use ordinal logistic regression to model categories of GAD-7 score to circumvent this violation of the homogeneity of variance assumption. Lastly, our study population did not include data from those of diverse backgrounds and the variables employed in our analyses are subject to selection bias and self-reporting bias. Hence, we urge that caution be taken when interpreting our results to any populations outside of the study sample.

Figures

	Quartile 1 (N=288)	Quartile 2 (N=295)	Quartile 3 (N=252)	Quartile 4 (N=333)	Overall (N=1168)
Baseline GAD-7 Score					
Mean (SD)	7.60 (4.84)	10.6 (5.06)	13.2 (4.90)	15.6 (4.88)	11.8 (5.77)
Median [Min, Max]	7.00 [0, 21.0]	10.0 [0, 21.0]	13.0 [3.00, 21.0]	17.0 [1.00, 21.0]	12.0 [0, 21.0]
6 Week GAD-7 Score					
Mean (SD)	6.77 (4.85)	8.91 (5.14)	11.2 (5.53)	12.8 (5.34)	9.97 (5.70)
Median [Min, Max]	6.00 [0, 21.0]	8.00 [0, 21.0]	11.0 [1.00, 21.0]	13.0 [0, 21.0]	9.00 [0, 21.0]
18 Week GAD-7 Score					
Mean (SD)	5.98 (4.45)	7.98 (4.81)	9.59 (5.20)	11.8 (5.77)	8.85 (5.53)
Median [Min, Max]	5.00 [0, 21.0]	7.00 [0, 21.0]	9.00 [0, 21.0]	11.0 [0, 21.0]	8.00 [0, 21.0]
12 Month GAD-7 Score					
Mean (SD)	6.39 (4.69)	7.97 (5.58)	9.20 (5.22)	9.59 (5.00)	8.33 (5.25)
Median [Min, Max]	5.00 [0, 19.0]	7.00 [0, 21.0]	8.50 [0, 20.0]	9.00 [0, 21.0]	7.00 [0, 21.0]
Sex					
Male	92 (31.9%)	73 (24.7%)	66 (26.2%)	62 (18.6%)	293 (25.1%)
Female	194 (67.4%)	221 (74.9%)	186 (73.8%)	268 (80.5%)	869 (74.4%)
Race					
White	226 (78.5%)	228 (77.3%)	198 (78.6%)	260 (78.1%)	912 (78.1%)
Black	31 (10.8%)	31 (10.5%)	24 (9.5%)	28 (8.4%)	114 (9.8%)
Other	31 (10.8%)	36 (12.2%)	30 (11.9%)	45 (13.5%)	142 (12.2%)
Education					
Less than High School	5 (1.7%)	4 (1.4%)	5 (2.0%)	7 (2.1%)	21 (1.8%)
High School	26 (9.0%)	27 (9.2%)	30 (11.9%)	51 (15.3%)	134 (11.5%)
Above High School	257 (89.2%)	264 (89.5%)	217 (86.1%)	275 (82.6%)	1013 (86.7%)

Table 1: Summary statistics for the total study population and by quartile of neuroticism score. Neuroticism score quartiles are defined as 0-29, 30-35, 36-39, 40-52 for quartiles 1, 2, 3, and 4, respectively. Quartiles are not equally sized due to the integer nature of neuroticism score. Continuous variables are summarized by Mean (SD) and Median [Min, Max] and categorical variables are summarized by N (%).

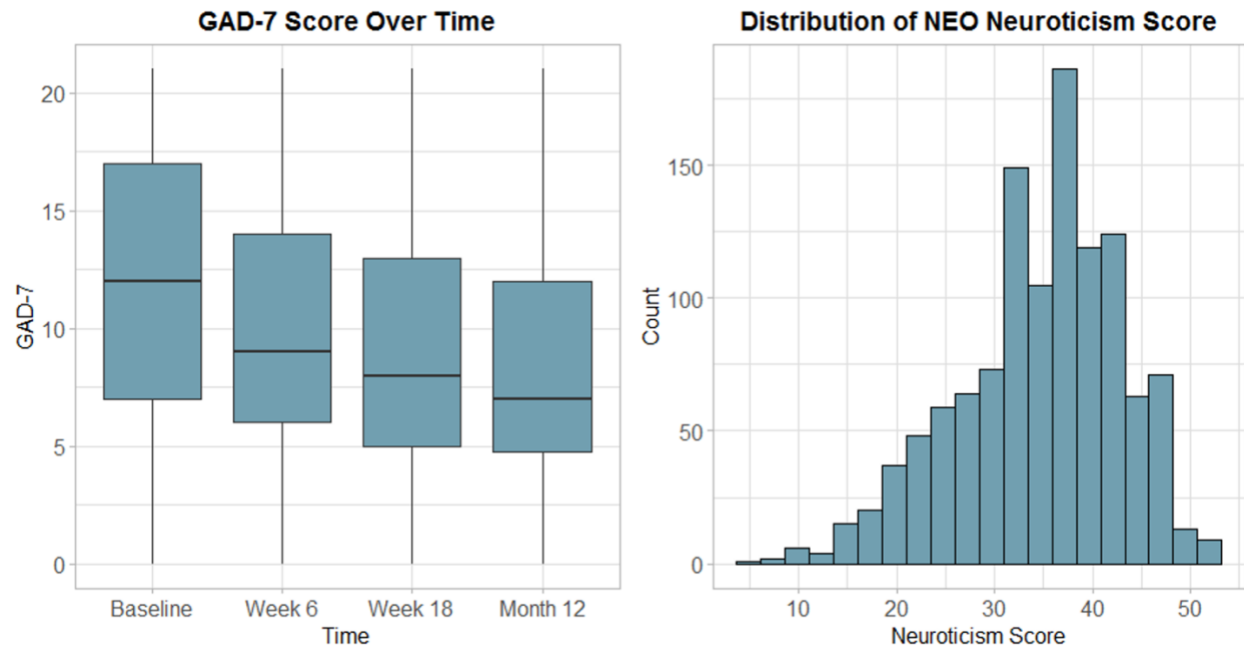


Figure 1: Distributions of GAD-7 score over time and neuroticism score

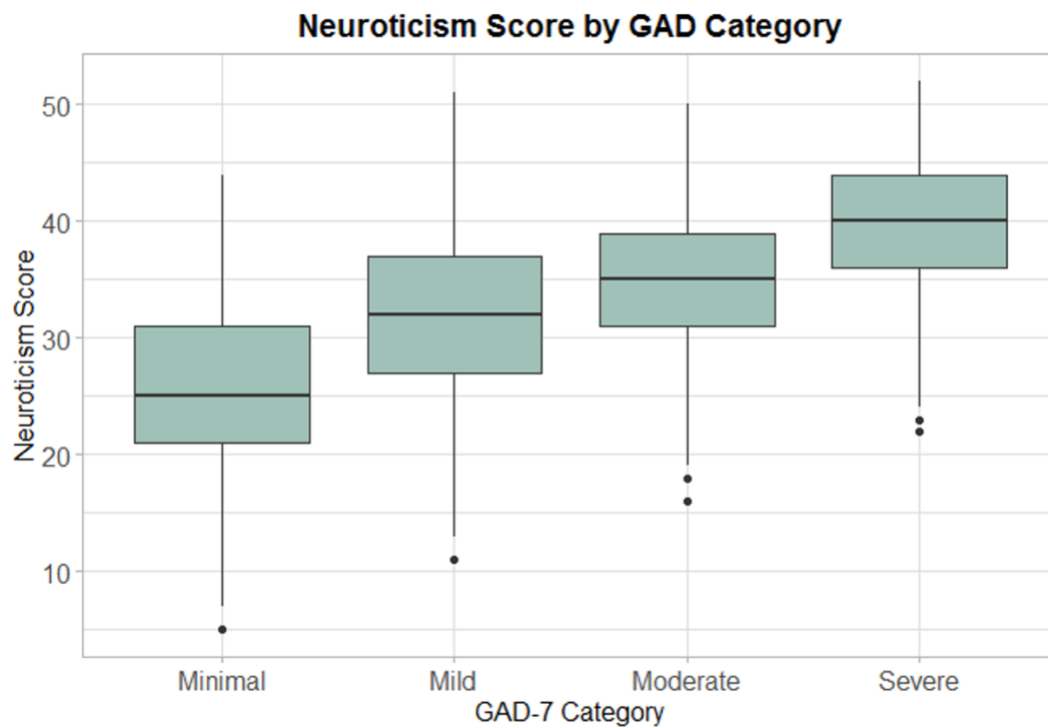


Figure 2: Distribution of neuroticism score for each category of GAD-7 scores, using baseline GAD-7 measurements

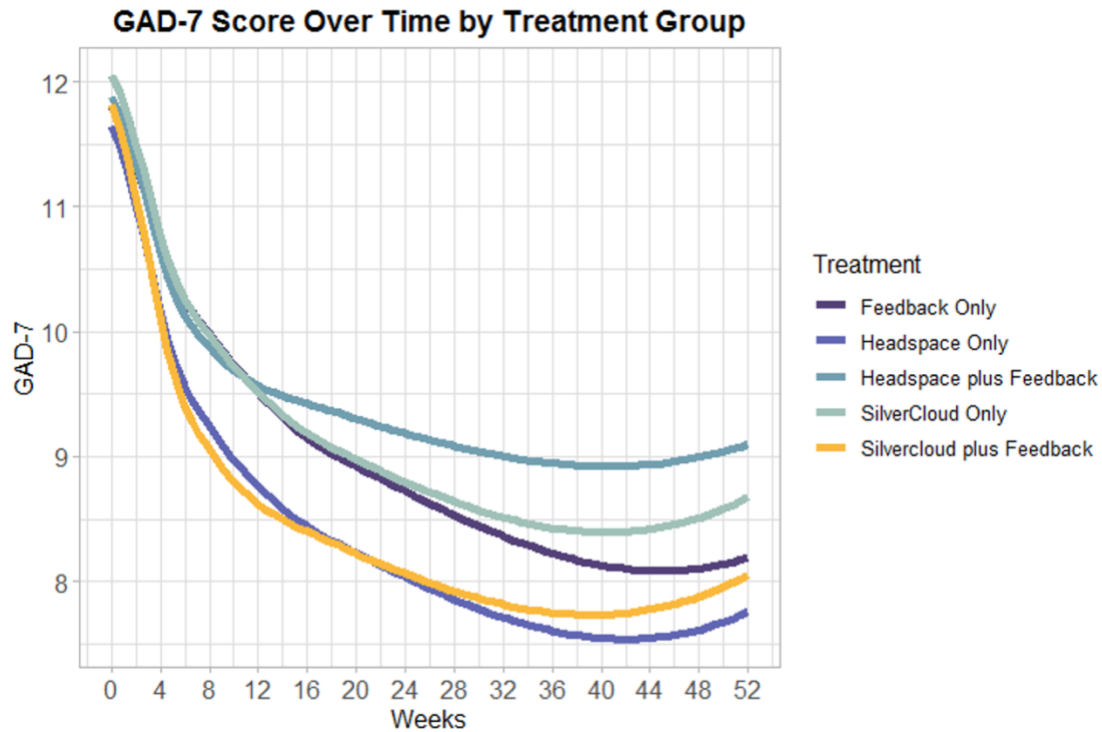


Figure 3: Loess smoothed curves of GAD-7 scores over time for each treatment group

<i>Predictors</i>	GAD-7 Score		
	<i>Incidence Rate Ratios</i>	<i>CI</i>	<i>p</i>
Time	0.86	0.84 – 0.87	<0.001
Neuroticism	1.35	1.31 – 1.39	<0.001
Headspace	0.94	0.87 – 1.02	0.140
Headspace plus Feedback	0.98	0.91 – 1.06	0.649
Silvercloud	1.02	0.95 – 1.10	0.595
Silvercloud plus Feedback	0.99	0.91 – 1.07	0.745
Age	0.96	0.93 – 0.99	0.003
Female	1.05	0.98 – 1.11	0.143
Black	1.00	0.92 – 1.09	0.996
Other Race	0.95	0.88 – 1.03	0.240
High School	0.99	0.81 – 1.22	0.941
Above High School	0.97	0.80 – 1.17	0.741
Time * Neuroticism	0.99	0.97 – 1.00	0.121

Figure 4: Parameter estimates from a model



Figure 5: Plot of residuals vs. fitted values for our final model

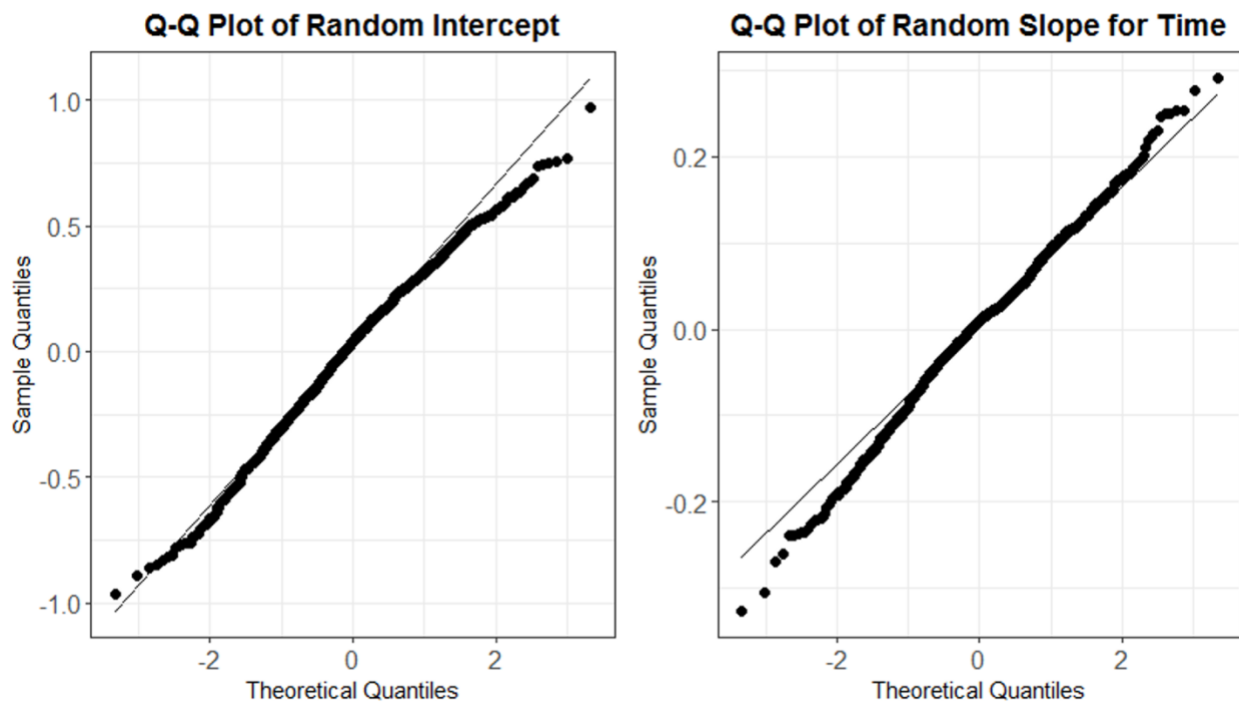


Figure 6: Quantile-quantile plots of random effects for final model to assess Normality of random effects

Appendix

<i>Predictors</i>	GAD-7 Score		
	<i>Incidence Rate Ratios</i>	<i>CI</i>	<i>p</i>
Time	0.86	0.84 – 0.87	<0.001
Neuroticism	1.37	1.31 – 1.44	<0.001
Headspace	0.94	0.87 – 1.02	0.139
Headspace plus Feedback	0.98	0.91 – 1.06	0.587
Silvercloud	1.03	0.95 – 1.11	0.520
Silvercloud plus Feedback	0.99	0.92 – 1.07	0.806
Age	0.96	0.93 – 0.99	0.003
Female	1.05	0.98 – 1.11	0.143
Black	1.00	0.91 – 1.09	0.977
Other Race	0.95	0.88 – 1.03	0.219
High School	0.98	0.79 – 1.20	0.812
Above High School	0.96	0.79 – 1.16	0.648
Time * Neuroticism	0.99	0.97 – 1.00	0.119
Headspace * Neuroticism	1.00	0.92 – 1.08	0.948
Headspace plus Feedback * Neuroticism	1.02	0.94 – 1.10	0.644
Silvercloud * Neuroticism	0.94	0.87 – 1.01	0.107
Silvercloud plus Feedback * Neuroticism	0.96	0.88 – 1.04	0.319

Supplement 1: Parameter estimates for a generalized linear mixed effects model modeling GAD-7 over time using neuroticism score, treatment group, age, sex, race, education, the interaction between time and neuroticism, and the interaction between treatment and neuroticism as predictors, including a random intercept and a random slope for time. Estimated model parameters for the interaction between treatment and neuroticism were not significant and were therefore taken out of the final model presented in this analysis.

$$Y_i \sim \text{Poisson}(\mu_i)$$

$$\begin{aligned} \log \mu_{ij} = & \beta_0 + \beta_1 \text{Time}_{ij} + \beta_2 \text{Neuroticism}_i + \beta_3 \text{Headspace}_i + \beta_4 \text{HeadspacePlusFeedback}_i \\ & + \beta_5 \text{Silvercloud}_i + \beta_6 \text{SilvercloudPlusFeedback}_i + \beta_7 \text{Age}_i + \beta_8 \text{Female}_i \\ & + \beta_9 \text{Black}_i + \beta_{10} \text{OtherRace}_i + \beta_{11} \text{HighSchool}_i + \beta_{12} \text{AboveHighSchool}_i \\ & + \beta_{13} \text{Time}_{ij} * \text{Neuroticism}_i + b_{0i} + b_{1i} \text{Time}_{ij} + \epsilon_{ij} \end{aligned}$$

Equation 1: Model equation for final generalized linear mixed effects model using a log link function

References

- New Study Reveals Lack of Access as Root Cause for Mental Health Crisis in America. (2018, October 10). *America's Mental Health 2018*. Retrieved from <https://www.cohenveteransnetwork.org/wp-content/uploads/2018/10/Press-Release-Americas-Mental-Health-2018-FINAL.pdf>.
- Rosellini, Anthony J., and Timothy A. Brown. "The NEO Five-Factor Inventory: Latent Structure and Relationships with Dimensions of Anxiety and Depressive Disorders in a Large Clinical Sample." *Assessment*, vol. 18, no. 1, 2010, pp. 27–38., <https://doi.org/10.1177/1073191110382848>.
- Yang, Xiaorong, et al. "Global, Regional and National Burden of Anxiety Disorders from 1990 to 2019: Results from the Global Burden of Disease Study 2019." *Epidemiology and Psychiatric Sciences*, vol. 30, 2021, <https://doi.org/10.1017/s2045796021000275>.