Tinnitus Project

STAT427 Project

2021-05-12

Consultants: Jesse Bowers, Shiwei Fu, Shuyu Jia, Yumeng Wu, Zhehui Chen

Client: Dr. Fatima Husain, Rafay Khan

Readme

Below we provide two versions of the report:

- Client Summary: This is a distilled version of the report which focuses on key findings.
- Main Body: This is a detailed version of the report which includes many more technical details. These details are not critical to understanding the findings, but provide a more comprehensive discussion of the methods and interpretation.

We recommend the Client Summary as the primary reference, as it is the most clear and concise. The main body and appendix can be referred to for additional information as needed.

In the addition to the report, we provide a Box link here. This folder includes the code used to generate our results. It also includes an appendix Excel with our EFA results. These files will be available here through 8/1/21.

Client Summary

Client Summary

Problem Statement

In this experiment, each test subject responded to five questionnaires: three tinnitus questionnaires (THI/TPFQ/THI), and two emotional questionnaires (BAI/BDI). We used their responses to answer two questions:

- Identify the best tinnitus questionnaire for clinical practice
- Identify how tinnitus influences emotion

We found the following results.

Data Description

Below we describe the data and call out any interesting patterns.

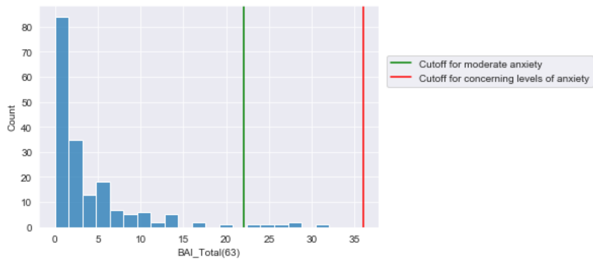
Demographics

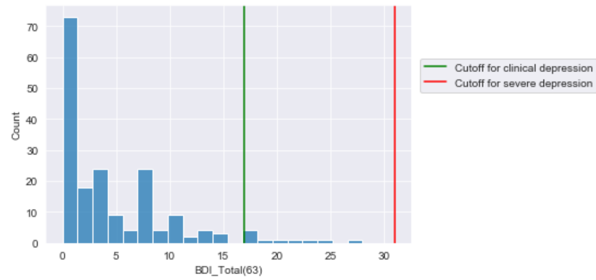
We found little influence of age, gender, and hearing loss onto tinnitus or emotion. Tinnitus and emotion are measured based on the total score of the survey responses.

	Tinnitus	**Emotion**
Age	Little Influence	Little Influence
Gender	Little Influence	Little Influence
Hearing Loss	Little Influence	**NEGATIVE** Influence

BAI/ BDI

We create frequency plots for the total score of BAI and BDA, which are both emotional measurements. We add two vertical lines to represent clinical thresholds.





For BAI (anxiety), we only have very few patients (6) that are considered as moderate anxiety and none of them have concerning levels of anxiety.

For BDA (depression), we only have very few patients (10) that are considered as clinical depression and none of them have severe depression.

As we will see later, having few patients at clinical levels of depression impacts our ability to predict emotion using tinnitus.

Milestone 1 – Validity

In this section we identify which tinnitus survey provides the most 'valid' measure of tinnitus. We use factor analysis to assess how well each tinnitus questionnaire performs on three measures:

- *Reproduction*: Can we reproduce the existing tinnitus subscales with EFA?
- *Internal Consistency*: How fully do the subscales explain responses to their own survey?
- *Comprehensiveness*: Which survey best explains the others and therefore is the most comprehensive?

EFA Consistency

Can we reproduce questionnaire subscales using EFA?

The following tables compare the existing subscales vs the reproduced subscales. The numbers represent the number of questions belonging to each subscale. The goal is to identify whether the reproduced EFA subscales and the existing subscales group questions in the same way.

THI:

The reproduced subscales group questions completely differently from the existing questionnaire.

Existing \ Reproduced EFA	Factor1: Severity	Factor2: Social	Factor3: Depression
Catastrophic	3	1	1
Emotional	4	2	2
Functional	8	4	0

EFA gives us one subscale measuring the overall severity of the tinnitus and one subscale measuring the social impact of tinnitus. For more information on how individual questions are categorized see the appendix EFA Excel.

TPFQ:

Here EFA largely reproduces the existing subscales:

Existing \ Reproduced EFA	F1: Sleep	F2: Hearing	F3: Concentration	F4: Emotion
S_Sleep	5	0	0	0
H_Hearing	0	5	0	0
C_Concentration	0	1	4	0

E_Emotion	0	0	2	3

Questions belong to the wrong subscale:

ID	Question	Sleep	Hearing	Concentration	Emotion	Existing subscales
3	I just wish my tinnitus would go away. It is so frustrating.	0.15	0.11	0.23	0.13	**Emotion**
5	When there are lots of things happening at once, my tinnitus interferes with my ability to attend to the most important thing.	0.16	0.34	0.24	0.13	**Concentration**
8	My tinnitus is annoying.	0.07	0.25	0.25	0.06	**Emotion**

The reproduced TPFQ is similar to the existing one. Questions 3, 5, 8 might be modified or removed from the questionnaire.

TFI:

Exploratory factor analysis puts almost all questions into the correct existing subscales. There are two exceptions:

ID	AUDITORY	COGNITIVE	SLEEP	RELAXATION	EMOTIONAL	QUALITY OF LIFE	INTRUSIVE	SENSE OF CONTROL
Q4	0.07	0.16	-0.16	0.3	0.06	-0.19	-0.06	0.29
Q22	-0.04	0.34	0.08	0.11	0.06	0.31	-0.03	0

In addition to questions Q4 and Q22, other questions belong to the same subscales as the existing TFI.

Key takeaways from EFA:

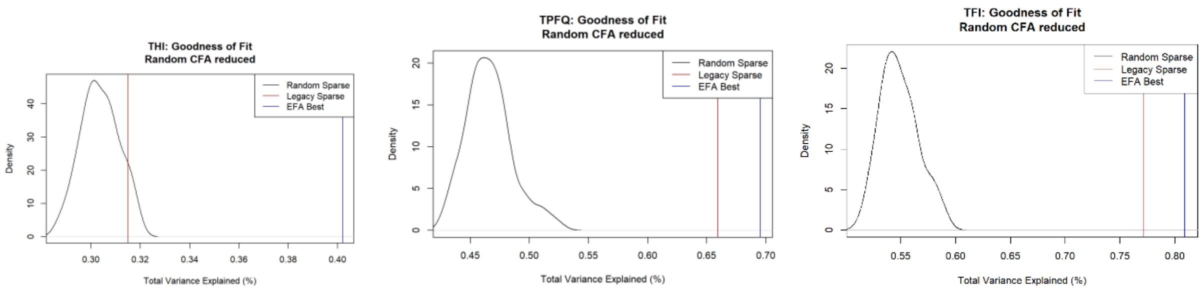
- THI was not reproduced by EFA
- TPFQ was mostly reproduced by EFA
- TFI was reproduced by EFA successfully

- TPFQ and TFI have a few questions which may be miscategorized. These questions might be removed to improve consistency.

Internal Validity

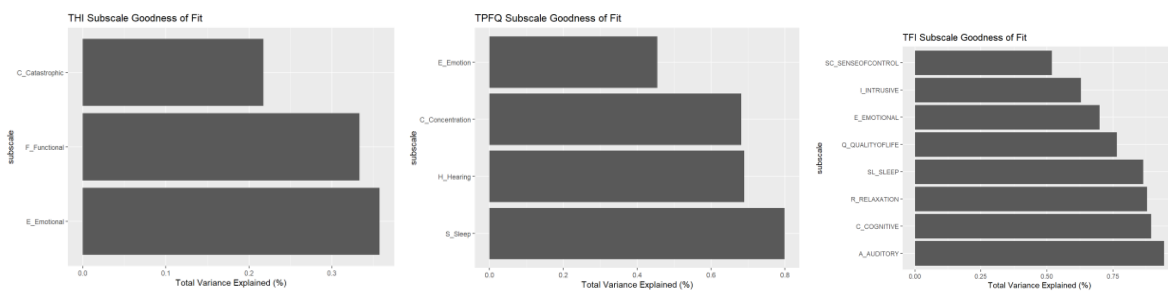
How well do the subscales describe the responses to that survey? Ideally if we knew the subscales of a survey, we would have a good idea of how that subject responded to individual questions in the survey.

First we examine how well the survey responses are explained given different choices of subscales:



Note: Above we examine how well different subscales explain the variation in survey responses across all questions. We compare the goodness of fit of the existing subscales (red), EFA custom subscales (blue), and randomly permuted subscales (black).

Now we examine how well each existing subscale explains the questions related to that subscale.



Note: Above we examine how well individual subscales explain their corresponding questions. For example how well does the emotional subscale explain variation in the emotion questions?

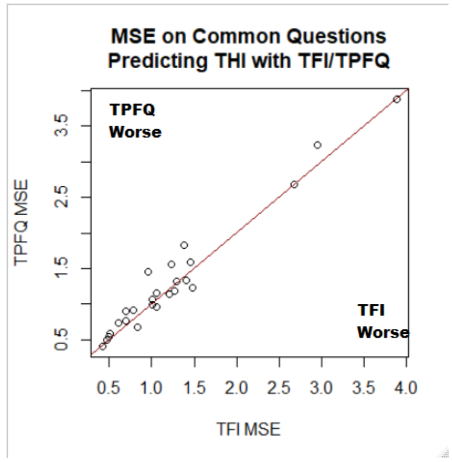
Key takeaways from internal validity:

- THI does little better than random subscales and much worse than ideal EFA.
- TPFQ and TFI do much better than random subscales and perform similarly to ideal EFA.
- TFI has slightly better internal validity, but this may be because it has more subscales allowing for a more customized fit.

Comprehensiveness

How well does one survey predict another survey? If one survey predicts another perfectly, then the first survey is comprehensive enough that there is no need to ask the second survey.

Overall Variation Explained (%)		Predicted Survey: THI	TPFQ	TFI
	THI		50%	47%
	TPFQ	61%		**63%**
Input	TFI	**65%**	**72%**	



Left: How well do the subscales of the input survey predict the subscales of the predicted survey, on average? e.g. The TFI subscales explain 65% of the variation in the THI subcales in a GLM.

Right: Each dot represents an individual THI question. That question is predicted with the TFI subscales (bottom) and the TPFQ subscales (right).

Predicting THI Explained Variance (%)

Predicted Subscale	from TPFQ	from TFI
THI-F_Functional	69%	70%
THI-E_Emotional	51%	61%
THI-C_Catastrophic	36%	36%

TPFQ predicting TFI

Predicted Subscale	Explained Variance (%)
TFI-SL_SLEEP	78%
TFI-C_COGNITIVE	71%
TFI-R_RELAXATION	66%
TFI-Q_QUALITYOFLIFE	62%
TFI-A_AUDITORY	61%
TFI-E_EMOTIONAL	56%
TFI-I_INTRUSIVE	51%
TFI-SC_SENSEOFCONTROL	48%

TFI Predicting TPFQ

Predicted Subscale	Explained Variance (%)
TPFQ-S_Sleep	78%
TPFQ-C_Concentration	74%
TPFQ-H_Hearing	68%
TPFQ-E_Emotion	67%

Note: How well do the subscales of one survey predict each individual subscale of another survey? Are there any subscales in survey B which are not well captured by survey A?

Key takeaways

- THI has a poor fit
- TPFQ and TFI produce similar quality fits when fitted to THI
- Neither TFI/TPFQ explain THI_Catastrophic well.
- Since TFI has more subscales, there are some subscales of TFI not explained well by TPFQ.

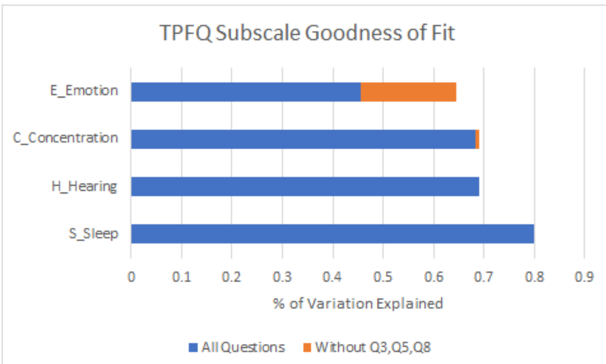
Removing Questions

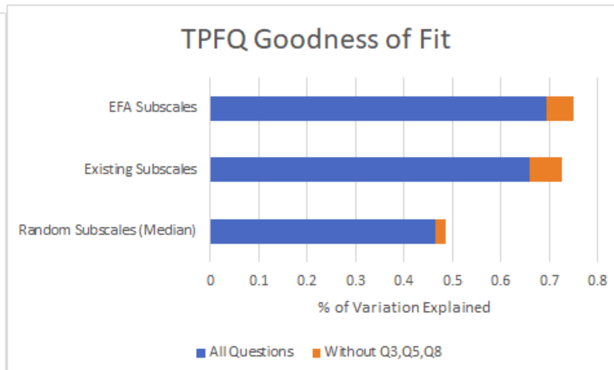
Based on our EFA analysis we identified certain questions with poor fit to their subscales:

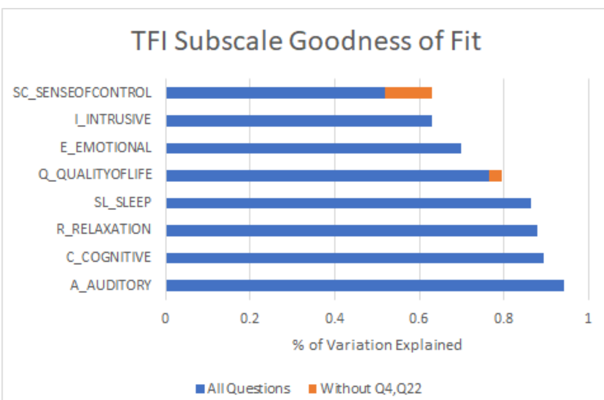
- TPFQ: Q3, Q5, Q8
- TFI: Q4, Q22

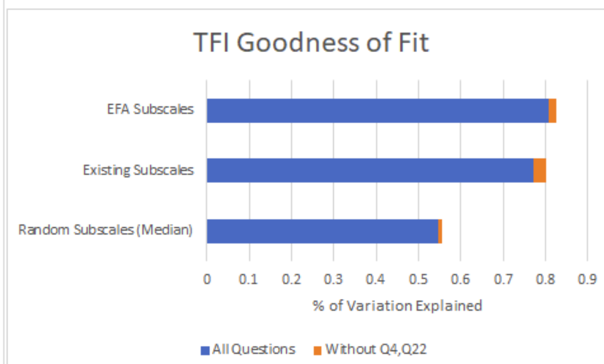
Below we examine the goodness of fit after removing those questions.

There is some data snooping here. We use the same sample both to identify the questions to remove and to measure the goodness of fit.









Left: We examine how well individual subscales explain the questions on that subscale. For example how well does the emotional subscale explain variation in the emotion questions?

Right: We examine how well different subscales explain the overall variation in survey responses.

Key takeaways:

- Removing questions improves goodness of fit particularly for the worst explained subscales.
  - Removing TPFQ Q3/Q8 seems to improve the Emotional subscale
  - Removing TFI Q4 seems to improve the Sense of Control Subscale
- The overall goodness of fit improves most for TPFQ

Milestone 1 – Conclusions

In this section we attempted to identify which tinnitus questionnaire gives the most 'valid' measure of tinnitus. In the EFA section, we identified whether we were able to reproduce the existing subscales with our data. In the internal validity section, we measured how completely the subscales explained responses to individual questions on that survey. In the comprehensiveness section, we measured how well one survey predicted another survey; the idea that if one survey predicted another well there is no need to ask the second survey.

The results were:

	EFA	**Internal Validity**	**Comprehensiveness**
THI	Not Reproduced	Poor Fit	Poor Fit
TPFQ	Reproduced, mostly	Good Fit	Good Fit
TFI	Reproduced, mostly	Good Fit	Good Fit

All three approaches agreed that the surveys from worst to best are THI < TPFQ < TFI. However the difference between TPFQ and TFI was consistently small, attributable at least in part to the fact that TFI has more subscales (overfitting). Removing certain questions from each survey seems to improve goodness of fit; especially for TPFQ.

The numbers give a slight advantage to TFI, but not enough to be conclusive. The choice comes down to this. Do you want a slightly more complicated measure (TFI) with more subscales or a slightly simpler/parsimonious measure (TPFQ) with fewer subscales? Both perform similarly.

Milestone 2 - Emotion

The goal for milestone 2 is to figure out whether it is possible to model the emotional disturbance from tinnitus based on the scores from BDI, BAI and the emotional subscales from the tinnitus questionnaires.

Part 1 Regression

While BAI and BDI certainly do not give us a comprehensive view of the true emotional state, they are likely to be the best response variables in our dataset. The predictors are emotional subscales from TFI, THI, and TPFQ. These subscales are calculated in percentages, therefore this is a regression problem. The results of building the model with the whole dataset is as follows (measured in mean absolute error):

Model	BAI	BDI
Linear Regression	3.29	3.61
Log Transform	3.10	3.45
Decision Tree	3.00	3.10
Random Forest	**1.94**	**2.10**
Gradient Boosting	**1.69**	**1.58**

Train-test split results:

Random Forests	BAI	BDI
Training MAE	1.99	1.64
Testing MAE	**3.47**	**5.20**

Gradient Boosting	BAI	BDI
Training MAE	1.74	1.69
Testing MAE	**2.50**	**4.13**

It is possible to model the emotional disturbance from tinnitus based on BDI, BAI and emotion subscales using techniques like Random Forests and Gradient Boosting, but the fit is poor. These models can be highly overfitting and are not useful to predict new observations. In addition, neither BAI score nor BDI score is good enough to be a response variable for this regression task, possibly due to the subjective nature of the self-report questionnaire.

Part 2 Classification

The regression on BAI and BDI did not give us a very good result. In this section we are going to try out the classification methods on BAI and BDI. To build up our classification methods, we started off from a very simple logistic regression model based on emotional subscales. Then we refined its responses and predictors to improve the classification accuracy. Finally, we applied up-sampling methods on BAI and BDI and corrected the skewness in the BAI and BDI dataset.

The best classification result is from this model:

BAI & BDI ~ 4 emotional subscales + other subscales in TFI

The “BAI & BDI'' identifies whether someone is emotionally disturbed by either BAI or BDI (or both) based on clinical cutoffs. We then applied up-sampling methods on this combined response to correct for the right skewness. Besides the 4 emotional subscales as we included in the regression part, we added any other subscales in TFI to make a higher classification accuracy.

The best classification accuracy after up-sampling is shown below:

BAI & BDI vs (3 Emotional Subscales + TFI)

Pred BAI & BDI	0	1
0	178	2
1	18	163

The classification accuracy is 0.89 and the No-Information-rate is 0.5 (after up-sampling), which is a good result. But there are several things that need to be considered before making a conclusion. Firstly, we applied up-sampling techniques, many of the observations in the data are not real. This good result might not be real. Moreover, we did model fitting and testing on the same data set, which might overestimate the prediction accuracy. Cross-validation is a way to fix this problem, but we do not have all real data at this moment, cross-validation on fabricated data still will not help much.

In conclusion, classification is a promising method to model the emotional disturbance. However, before we make a solid conclusion that classification is for sure to work, we need to collect more data, especially a considerably larger group of patients who would be classified as emotionally dsturbed.

Conclusions

Based on our analysis we conclude the following.

Milestone 1 - Choosing the Best Survey

- We recommend TFI as the primary tool to measure Tinnitus. To further improve validity you could remove questions Q4 and Q22.
- TPFQ is also a good candidate. It's validity is slightly worse, but this is due largely to having fewer subscales. TPFQ validity is improved significantly by removing questions Q3, Q5, and Q8.
- THI has poor fit relative to the other surveys.

Milestone 2 - Influence of Tinnitus onto Emotion

- Neither BAI nor BDI is good enough to be a response variable for regression, possibly due to the subjective nature of the self-report questionnaire.
- Ensemble methods for regression overfit and offer poor predictions.
- Using median as a cutoff, classification on BAI and BDI performs poorly with a simple logistic model.
- After up-sampling, classification fits well. This allows us to use the scientific cutoffs of each questionnaire even though our data is unbalanced.

Main Body

Main Body

Background

Problem Statement

In this experiment, each test subject responded to five questionnaires: three tinnitus questionnaires (THI/TPFQ/THI), and two emotional questionnaires (BAI/BDI). We used their responses to answer two questions:

- Identify the best tinnitus questionnaire for clinical practice
- Identify how tinnitus influences emotion

We found the following results.

Definition of Tinnitus

Tinnitus is when you experience ringing or other noises in one or both of your ears. The noise you hear when you have tinnitus isn't caused by an external sound, and other people usually can't hear it. Tinnitus is a common problem and analyzing the potential factors that are associated with tinnitus is important for clinical practices.

Questionnaires Data

- Tinnitus Handicap Index (THI)
- Tinnitus Functional Index (TFI)
- Tinnitus Primary Function Questionnaire (TPFQ)
- Beck Anxiety Inventory(BAI)
- Beck Depression Inventory (BDI)

The datasets are collected by the clients and they are in the questionnaire format. The five questionnaire datasets we will use are THI, TFI, TPFQ, BAI and BDI. The first three questionnaires assess tinnitus severity and they measure different subscales. Subscale means different measurements that contribute to tinnitus. The last two questionnaires are all self-reported emotional reactions, which measure patients' emotional problem and they do not have any subscales.The questionnaire together with its subscales are shown below.

Tinnitus Handicap Index (THI) -- Functional, Emotional, Catastrophic.
Tinnitus Functional Index (TFI) -- Intrusive, Sense of Control, Cognitive, Sleep, Auditory, Relaxation, Quality of life, Emotional.
Tinnitus Primary Function Questionnaire (TPFQ) -- Sleep, Emotion, Concentration, Hearing.

For more detailed information about which questions measure different types of subscales, please refer to the appendix.

Combined Dataset

Questionnaires					THI				TPFQ					TFI									BAI	BDI-II
Study	Location	Hearing status	Age	Gender	Total/100	(F) %	(E) %	(C) %	Total %	(C) %	(E) %	(H) %	(S)%	Total/100	(I)%	(SC)%	(C)%	(SL)%	(A)%	(R)%	(Q)%	(E)%	Total/63	Total/60
A	UIUC	HL	52	F	84.00	91.67	87.50	60.00	89.70	86.80	91.00	92.00	89.00	81.60	95.00	90.00	75.00	86.67	88.33	75.00	65.00	83.33	16.00	18.00
A	UIUC	HL	60	F	8.00	0.00	0.00	40.00	5.05	0.20	20.00	0.00	0.00	11.20	40.00	30.00	10.00	10.00	3.33	0.00	0.00	0.00	0.00	1.00
A	UIUC	NH	40	M	6.00	4.17	0.00	20.00	1.60	3.40	2.40	0.40	0.20	6.00	10.00	30.00	0.00	0.00	10.00	0.00	0.00	0.00	0.00	1.00
A	UIUC	HL	47	M	18.00	20.83	12.50	20.00	31.25	20.00	43.00	20.00	42.00	29.20	70.00	66.67	16.67	10.00	23.33	40.00	10.00	3.33	1.00	0.00
A	UIUC	HL	45	M	12.00	8.33	6.25	30.00	10.00	10.00	6.00	0.00	24.00	8.00	20.00	16.67	3.33	3.33	0.00	20.00	0.00	3.33	0.00	1.00
A	UIUC	HL	57	M	42.00	54.17	25.00	40.00	71.25	76.00	74.00	76.00	59.00	55.20	36.67	56.67	53.33	73.33	70.00	66.67	47.50	40.00	3.00	9.00
A	UIUC	HL	58	F	62.00	54.17	75.00	60.00	68.75	87.00	61.00	57.00	70.00	79.20	76.67	83.33	80.00	76.67	80.00	80.00	80.00	76.67	11.00	0.00
A	UIUC	HL	49	M	10.00	12.50	0.00	20.00	7.75	7.00	9.00	15.00	0.00	6.80	20.00	13.33	0.00	0.00	20.00	0.00	2.50	0.00	3.00	0.00
A	UIUC	HL	53	M	46.00	54.17	31.25	50.00	60.50	70.00	58.00	52.00	62.00	66.80	90.00	76.67	80.00	50.00	80.00	43.33	52.50	66.67	11.00	15.00
A	UIUC	NH	25	M	2.00	4.17	0.00	0.00	0.75	0.00	2.00	0.00	1.00	2.80	20.00	0.00	3.33	0.00	0.00	0.00	0.00	0.00	3.00	9.00
A	UIUC	HL	47	M	12.00	12.50	6.25	20.00	7.00	9.00	16.00	3.00	0.00	8.40	36.67	6.67	0.00	0.00	20.00	0.00	2.50	3.33	1.00	0.00

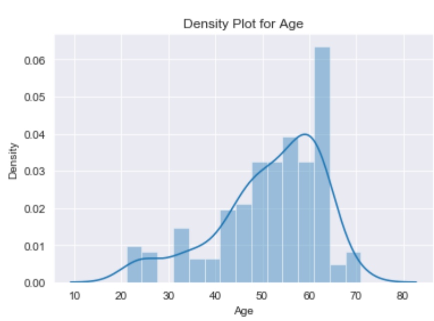
The combined dataset has 190 observations with 3 observations that contain missing values. The combined dataset contains three main parts. The first part is questionnaire information, including study type, location and hearing status. The second part is demographic features, including age and gender. The rest columns are all scores calculated based on the responses from all the patients using different formulas. For the three total scores that measure tinnitus severity, the total scores are in a scale of 0-100 and all the scores of the subscales are in percentage format. The other two scores that measure emotional problems are in a range of 0-63.

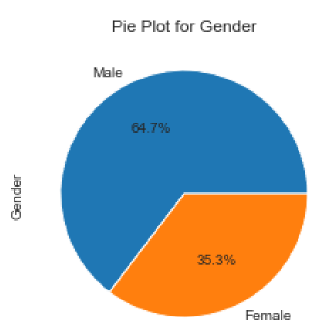
Data

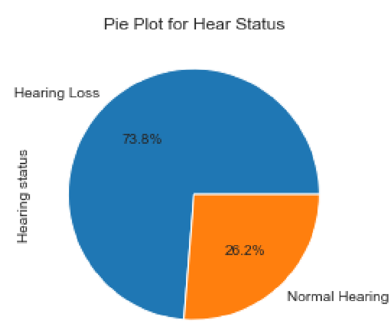
We merged responses across questionnaires and dropped any observations with missing responses.

Demographics

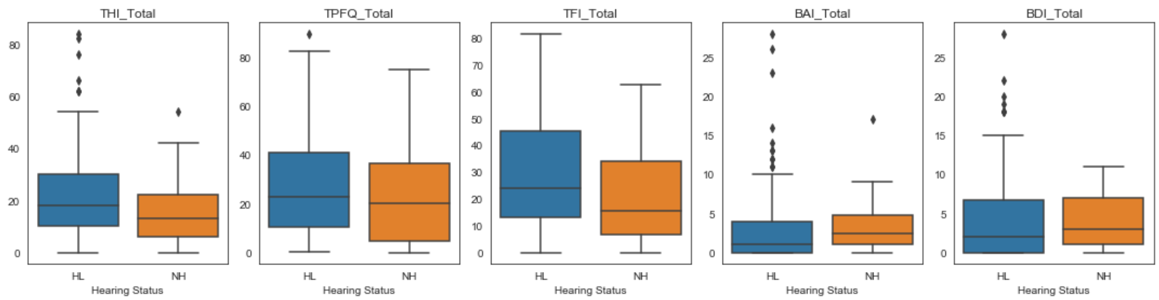
For the three studies, the protocol for recruiting subjects to adults who had tinnitus and did not have severe emotional problems. Therefore the demographic makeup is quite a bit different from the average US population. We have a higher frequency of older individuals, men, and people with hearing loss.

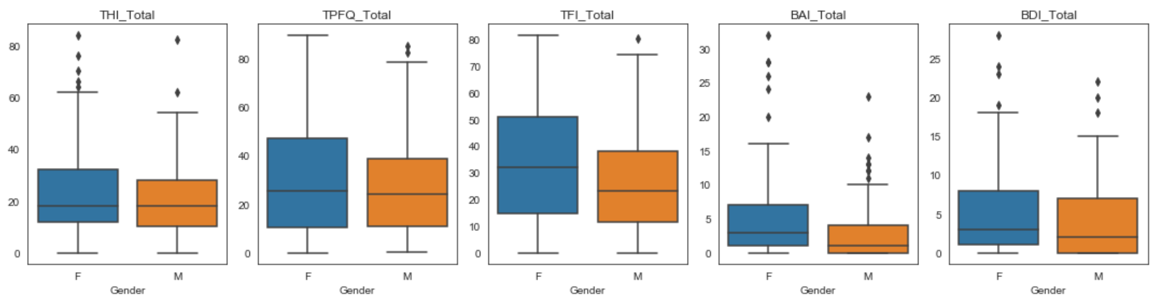


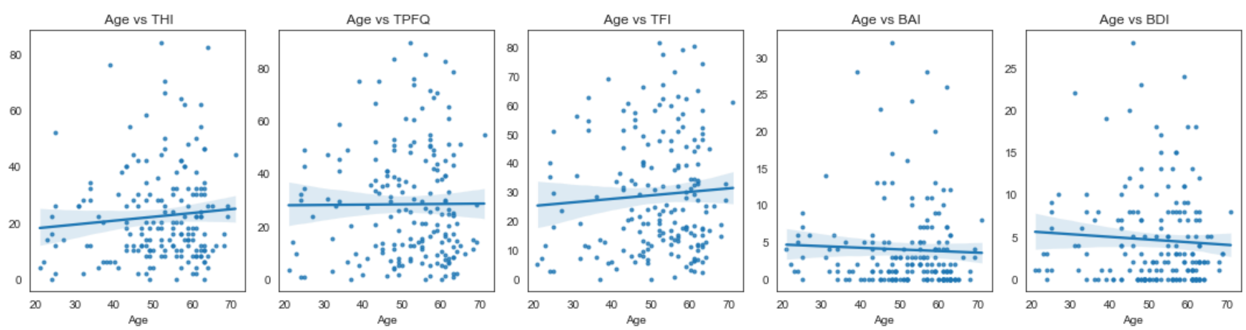




Our analysis did not show any strong relationship between these demographics and the tinnitus responses (below). You can see below that generally people without hearing loss tend to have higher emotional scores indicating that when there is no clear cause for the tinnitus there may be some emotional component. Furthermore, females have slightly higher tinnitus and emotional total scores. There is no strong age relationship.



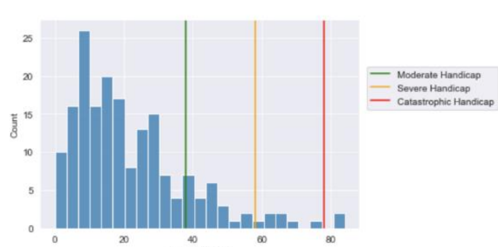


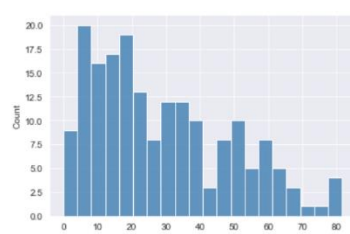


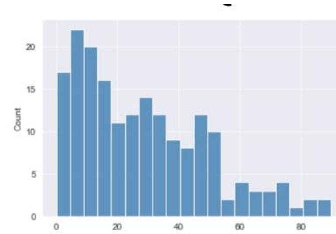
In general our analysis did not find any strong relationships between demographics and tinnitus. Therefore we do not include demographic interactions in our analysis for milestone 1.

THI/ TPFQ/ TFI

We create frequency plots for the total score of THI, TFI and TPFQ, which are all tinnitus severity measurements and contain different subscales. The distributions of the three total scores are right skewed and very few observations that lie beyond the severity cutoff.

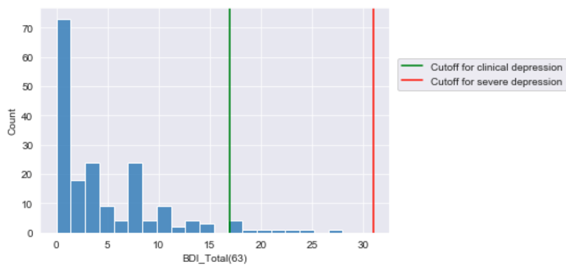


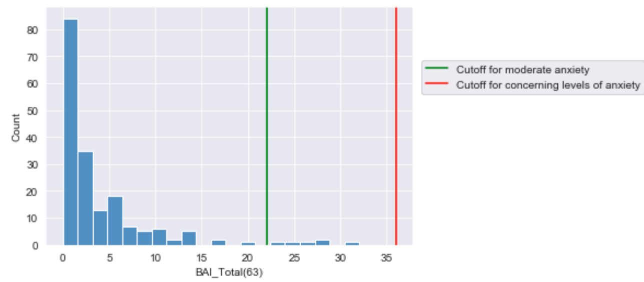




BAI/ BDI

We create frequency plots for the total score of BAI and BDI, which are all emotional measurements. We add two vertical lines to represent the thresholds.





For BAI (anxiety), we only have very few patients (6) that are considered as moderate anxiety and none of them have concerning levels of anxiety.

For BDA (depression), we only have very few patients (10) that are considered as clinical depression and none of them have severe depression.

	Tinnitus	**Emotion**
Age	Little Influence	Little Influence
Gender	Little Influence	Little Influence
Hearing Loss	Little Influence	**NEGATIVE** Influence

Conclusions of Exploratory Data Analysis

If we look at relationships, age, gender and hearing loss have little influence on patients' annoyance generated by tinnitus. While hearing loss tends to have negative impacts on people's emotions.

If we look at distributions, Only a small portion of the patients actually suffer from tinnitus or emotional problems.

Factor Analysis - Background

A survey might have a large number of questions, but only ask about a few key ideas. For example a 20 question math exam might measure someone's ability to add and multiply; with several questions on either topic. A big five personality test might have 60 questions, but only measures five personality traits: openness, conscientiousness, extraversion, agreeableness, and neuroticism. In either case it's not efficient to look at each question individually. Instead we want to group similar questions into subscales, look across questions, and evaluate these larger ideas.

Factor analysis is a statistical tool for building and evaluating groups/subscales. Exploratory factor analysis (EFA) is a way to 'explore' the data and identify new groups/subscales. Confirmatory factor analysis (CFA) is a way to 'confirm' the quality of existing subscales. We assume these subscales are measuring some real (but unobserved) phenomena (e.g. addition problems measure skill with addition). We call these phenomena latent factors.

Traditional factor analysis assumes continuous normal responses to each question. Our tinnitus surveys are not continuous. Instead the subject is given a small set of choices (e.g. 0/2/4) ordered from smallest to largest (i.e. likert/ordinal responses). There are three main approaches for adapting factor analysis to ordinal data.

<u>Structural Equation Modeling (SEM)</u>: We assume that the ordinal responses are the result of rounding. Underlying each response is an actual continuous number which is rounded to the closest value (e.g. 0/2/4). The latent factors influence these continuous responses which in turn determine our rounded responses.

<u>Item Response Theory</u>: We treat our responses entirely as discrete. But we allow for latent factors to influence the probability of choosing a given response. Some examples are the Graded Response Model (Samejima 1997), and the MIRT (Chalmers 2012).

<u>Correlation Based Methods</u>: We pretend the data is continuous. We build a correlation matrix describing the relationship between questions. Then we pretend those associations were generated from a normal random variable instead of our ordinal responses. The correlation matrix could be built using usual (pearson) correlations, polychoric correlations, or spearman rank correlations. Polychoric and spearman correlations recognize the responses are ordinal, and make adjustments to get a more accurate correlation. For example see LISREL (Joreskog 1994), and Lavaan (Igolkina 2020 appendix C).

IRT/SEM are robust and interpretable from a statistical perspective. In comparison the correlation method impersonates continuity and lacks the same rigor. However the more formal IRT/SEM methods require many more parameters compared to the correlation method. Additionally there is less software support for IRT/SEM (particularly for CFA). Due to our small sample size and software needs, we opted for a correlation based method using Spearman Correlation. The tradeoff is we cannot reverse the continuity assumption. We cannot take this continuous random variable, convert it back into the discrete response, and compare our estimates against the actual responses. In particular we cannot calculate proper probabilities or likelihoods.

In general our factor model is:

Response = Loadings * Factors + Noise

$$X = \Lambda F + E$$

Where factors (F) are the hidden underlying phenomenon we are studying, and the loadings (Λ) describe how strongly a factor relates to a given individual question.

Milestone 1 - Methods

In milestone 1, we restricted our analysis to subjects which completely filled out all of the tinnitus surveys. This made sure our results were consistent across all of our comparisons. Our analysis proceeded as follows:

1. Internal Validity: We examined each questionnaire individually and measured its validity.

1A. Reliability: We measure to what extent the questions in a survey support the building of one or more subscales. We looked at several metrics for this: Cronbach's Alpha, EFA, and Bartlett Sphere test, and KMO. We considered a variety of quality measures for EFA since the Spearman Correlation does not permit a proper calculation of likelihood, AIC, or BIC due to non-normality. Instead, we used cumulative variance as the metric. The cumulative variance represents the proportion of the total variance caused by all common factors.

Also, we used KMO score and Bartlett's Sphericity test to check the suitability of the dataset for factor analysis . The KMO value represents the simple correlation coefficient and partial correlation coefficient between the comparison variables. The value range of KMO is 0-1. If the sum of squares of simple correlation coefficients is much larger than the sum of squares of partial correlation coefficients, and the KMO value is close to 1. The stronger the correlation between variables, the more suitable for factor analysis. The Bartlett sphere test is a coefficient to test whether the variables are correlated. If the p-value is less than 0.05, it means that the variables are correlated and suitable for factor analysis.

Cronbach's Alpha is a measure of internal consistency and the reliability of the scale. When we use multiple items to measure a latent variable, we call this set of items a scale. For example, we want to know whether Tinnitus makes people feel stressed. If we ask this question directly, we will not get reliable results because it is up to the patients to define what is feeling. Alternatively, we can construct a scale such as "Yes", "Sometimes" and "No", or a scale of "Always", "Sometimes", "Rarely", and "Never", or a scale of 1-5. To test whether these measures are useful and consistent, we use Cronbach's Alpha. Generally, a Cronbach's Alpha of .7 or more is considered acceptable.

1B. CFA: We used confirmatory factor analysis (CFA) to take our existing subscales, fit an analogous model, and evaluate the goodness of fit. We focused on 'sparse' CFA which equally weights all questions in a subscale (only 0/1's as loadings). We opted for a limited information method using Spearman rank correlation. Spearman rank correlation uses the same formal as traditional (pearson) correlation, but does correlation on inferred rankings rather than raw values. In a limited information method our data is used to generate an artificial correlation method, and that artificial correlation method is fit under CFA. Because our CFA is fit to an artificial construct rather than directly on our data, we cannot calculate likelihoods or measures derived from likelihoods such as AIC/BIC. Instead we evaluate goodness of fit based on the percent of Spearman correlation explained. Remember that Spearman correlation is an artificial construct, and we are using it as a proxy to estimate goodness of fit.

1C. EFA: We use EFA to build custom subscales fitted directly on our data (no sparse 0/1 loadings assumed). We then interpret the results and see if they are consistent with the legacy subscales. We calculated spearman correlation due to the ordinal data (Likert scale) and used maximum likelihood method to find the parameters. To prevent overfitting and have a better one to one comparison with the existing subscales, we use the legacy amount of factors.

2. External Validity: We examine how well one questionnaire explains another questionnaire. The most comprehensive questionnaire should explain the others best.

2A. Survey to Survey: We examine how well do the subscales from one survey explain all of the surveys from another survey. We fit multiple linear regression predicting each individual subscale, and then look at the average reduction in variation.

2B. Survey to Subscale: We examine how well do the subscales from one survey explain an individual subscale from another survey.

2C. Survey to Question: We examine how well do the subscales from one survey explain individual questions from another survey. Since the questions are ordinal we use a proportional odds model. Since different surveys have different scales (0,2,4 vs 0-100) we grouped the responses of each survey into three (0, 1-50, 51-100). We evaluated the accuracy of our models with mean squared error (MSE).

Milestone 1 – EFA Consistency

We attempted to reproduce the legacy subscales organically. We would build new subscales on our data using EFA and then compare our fitted models against the legacy subscales. If we are able to reproduce the existing subscales this validates the choice of subscales. Three kinds of rotation methods were used on factor loadings to get more interpretable and reliable results, including varimax, oblimin, and target rotation. Varimax is an orthogonal rotation method that assumes the common factors are uncorrelated. In contrast, oblimin is an oblique rotation method that assumes that the factors are correlated. Target rotation is an intermediate process between CFA and EFA. This method tests a specific hypothesis for the common factors.

Considering the subscales are correlated with each other, we analyze the factor loadings after oblimin rotation. The details of factor loadings are in the excel file. We only show the factor loadings that are greater than 0.2. Factor loadings less than 0.2 are hidden, but you can see them by changing the text color.

For THI our fitted models were inconsistent with the existing subscales. For TPFQ we had a similar result. For TFI we were able to closely match the existing subscales using EFA.

THI

Existing \ Reproduced EFA	Factor1: Severity	Factor2: Social	Factor3: Depression
Catastrophic	3	1	1
Emotional	4	2	2
Functional	8	4	0

The reproduced THI has 3 common factors which only explain 40% of the total variance together (cumulative variance). Factor1 represents the overall severity. Factor2 separates out the social impacts. Factor3 separates out depression from other emotions. The contingency table shows whether the questions in reproduced THI are mislabelled as the wrong subscale or not. We can see that we got a completely different result from the existing THI questionnaire.

TPFQ

ID	Question	Sleep	Hearing	Concentration	Emotion	Existing subscales
3	3 I just wish my tinnitus would go away. It is so frustrating.	0.15	0.11	0.23	0.13	**Emotion**
5	5 When there are lots of things happening at once, my tinnitus interferes with my ability to attend to the most important thing.	0.16	0.34	0.24	0.13	**Concentration**
8	8 My tinnitus is annoying.	0.07	0.25	0.25	0.06	**Emotion**

Existing \ Reproduced EFA	F1: Sleep	F2: Hearing	F3: Concentration	F4: Emotion
S_Sleep	5	0	0	0
H_Hearing	0	5	0	0
C_Concentration	0	1	4	0
E_Emotion	0	0	2	3

TPFQ has 20 items assessing 4 primary subscales: Emotion, Hearing, Sleep and Concentration. The cumulative variance of TPFQ is 0.70, which is acceptable. The factor loadings of Question 3, 5, 8 are smaller than 0.4, which makes it hard to find the common factors that they belong to. The contingency table shows that these 3 questions are not assigned to the right subscales in existing TPFQ. Therefore, these questions need to be modified or removed from the questionnaire. To improve the structure of the TPFQ questionnaire, we deleted these 3 questions and rebuilt the EFA. After deleting questions, the cumulative variance changes to 0.75. Therefore, we can delete Question 3, 5, 8 to make the structure more concise.

TFI

ID	AUDITORY	COGNITIVE	SLEEP	RELAXATION	EMOTIONAL	QUALITYOFLIFE	INTRUSIVE	SENSEOFCONTROL
Q4	0.07	0.16	-0.16	0.3	0.06	-0.19	-0.06	0.29
Q22	-0.04	0.34	0.08	0.11	0.06	0.31	-0.03	0

The reproduced TFI has almost the same structure as the existing questionnaire. Only Question 4 and 22 belong to the wrong subscale. We deleted these 2 questions and rebuilt the EFA model. After deleting questions, the cumulative variance changed from 0.81 to 0.84. It makes sense to delete Question 4 and 22.

Conclusions

THI was not reproduced by EFA so it's not a good choice for statistical analysis. We did not rebuild the EFA and CFA models for THI due to the huge difference between the reproduced results and the existing results. TPFQ was mostly reproduced by EFA and the result was acceptable. However, some emotion questions were assigned to the concentration subscale. TFI was reproduced by EFA successfully. After deleting these mislabelled questions, the cumulative variance of TPFQ and TFI both increased.

In general, TFI is the best one based on exploratory factor analysis, but it may have overfitting problems. TPFQ is acceptable. It's best not to use THI for research.

Milestone 1 – Internal Validity

In this section we examined each questionnaire individually and measured its validity.

We will show that TFI and TPFQ perform similarly with TFI having a modest advantage.

Question Strength

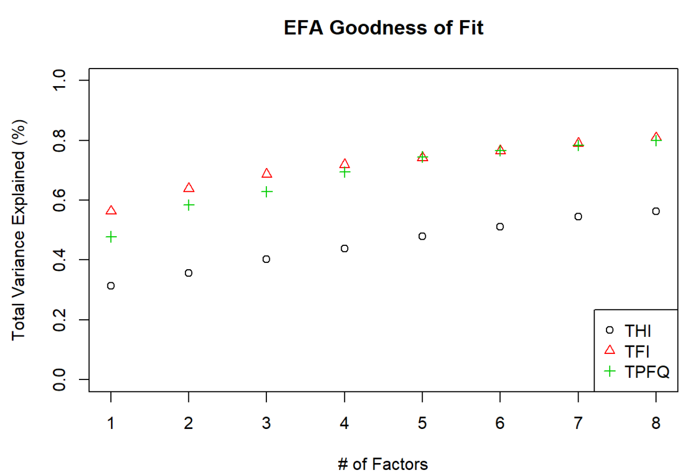
In this section we measure to what extent the questions in a survey support the building of one or more subscales.

Reliability: Are the questions in a survey interrelated? Ideally we want a survey to focus on a few main topics with questions on each topic closely related. Cronbach's alpha is one way to measure the degree of relationship on a scale from zero to one. One of the strengths of Cronbach's alpha is we can measure without needing to specify specific topics, number of topics, or which questions should be grouped together.

	THI	TPFQ	TFI
Cronbach's alpha	0.92	0.95	0.97

Above we can see that all questionnaires have strong relationships (>0.8) with TFI the strongest by a small margin.

Dimensionality: If we used the same number of factors which questionnaire would perform best? Below we use EFA to build new subscales of different sizes. For each questionnaire we start to build between 1 and 8 total subscales, and measure the goodness of fit.



Above we can see that for any fixed number of factors that TFI and TPFQ perform similarly (TFI slightly better), and THI performs the worst.

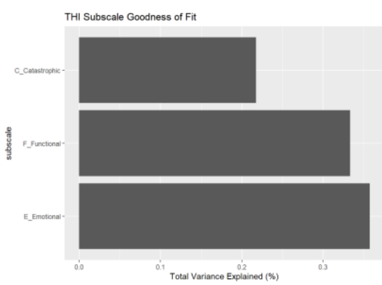
Other Measures: We also used the Bartlette's Sphericity test and KMO score to evaluate the appropriateness of subscales (and factor analysis). Consistent with the above we find evidence for subscales with TFI performing best. See definitions above (#XMeasures).

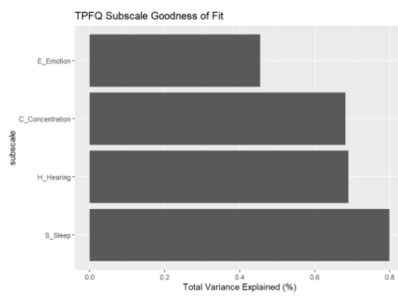
	Good If	THI	TPFQ	TFI
Bartlett Sphere P-Value	<5%	<1%	<1%	<1%
KMO	>0.8	0.88	0.91	0.94

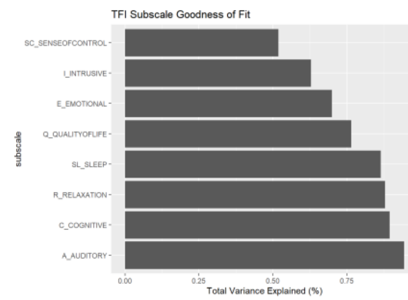
Subscale Accuracy with CFA:

How well do the subscales describe the responses to that survey? Ideally if we knew the subscales of a survey, we would have a good idea of how that subject responded to individual questions in the survey.

Individual Subscales: How well does each subscale explain its corresponding questions? Some subscales perform better than others. Below we measure how much of the spearman correlation of each question group is explained by its corresponding subscale.

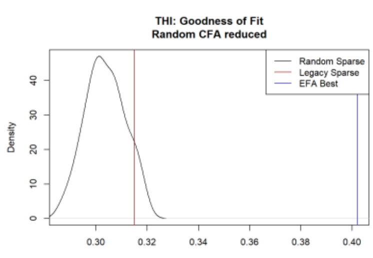


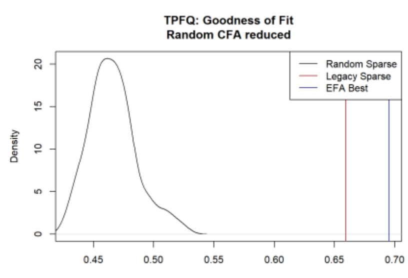


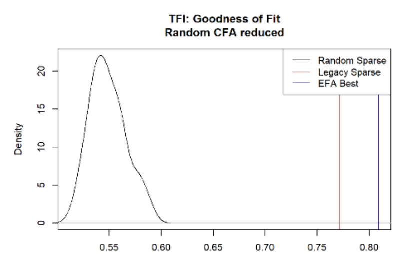


Above we see that the THI subscales provide the worst explanations. TPFQ and TFI have roughly similar goodness of fit. Note that THI_Catastrophic and TPFQ_Emotion have the worst reliability.

Overall Quality: How well do the subscales combined explain the responses? We use confirmatory factor analysis to fit the legacy subscales to the observed spearman correlation. We then offer two comparison points. As a 'worst case' comparison, we generate 100 random subscales. As a 'best case' comparison, we fit an EFA model which builds a subscale specifically customized to our data. We expect the legacy subscales to sit somewhere between either extreme.







Total Variance Explained (%)	THI	TPFQ	TFI
Lower Bound: Random Subscale (95%)	[0.289, 0.317]	[0.433, 0.511]	[0.521, 0.585]
Legacy Subscale	**0.315**	**0.659**	**0.772**
Upper Bound: EFA Subscale	0.402	0.695	0.809
Subscales (#)	3	4	8

Above we see that THI by far performs the worst; little better than random subscales. TPFQ and TFI both perform well with TFI as the best. Note that the TFI legacy subscales perform better than a five factor EFA model. Additionally TFI surpasses random subscales by a wider margin compared to TPFQ.

Other Measures: Below are some additional measures of the goodness of fit of the provided subscales. These results are consistent with our analysis above. TFI subscales perform the best and THI subscales often perform the worst.

	Cut-off for good fit	THI	TPFQ	TFI
RMSEA	RMSEA < 0.08	0.092	0.130	0.074
SRMR	SRMR <0.08	0.102	0.092	0.094
CFI	CFI ≥0.90	0.736	0.840	0.950
TLI	TLI≥ 0.95	0.730	0.831	0.944

For more information about these measures see the appendix (#XMeasures).

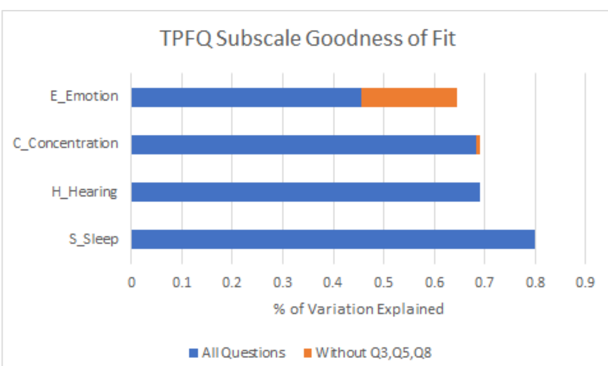
Removing Questions

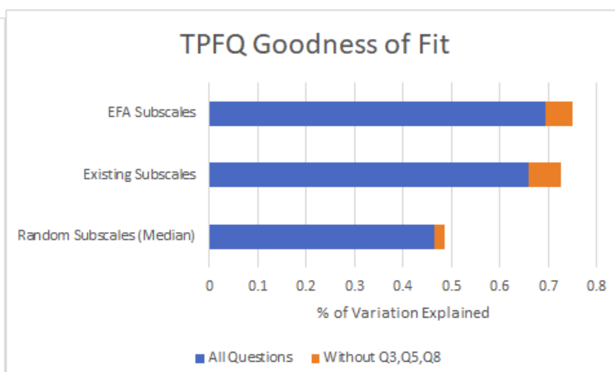
Based on our EFA analysis we identified certain questions with poor fit to their subscales:

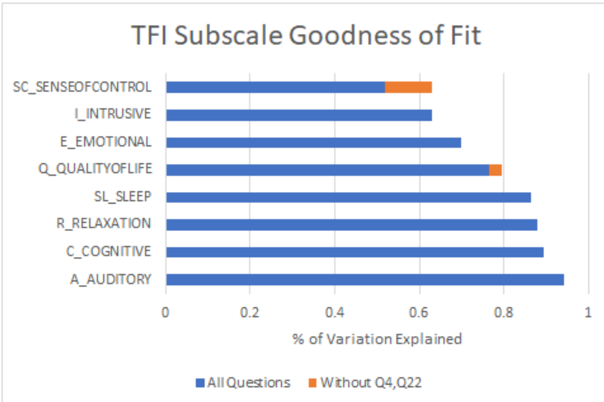
- TPFQ: Q3, Q5, Q8
- TFI: Q4, Q22

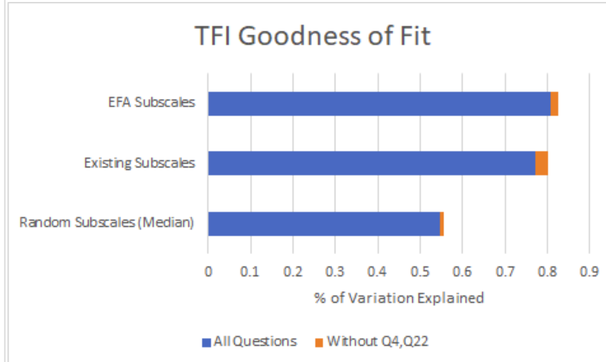
Below we examine the goodness of fit after removing those questions.

There is some data snooping here. We both identify the questions to remove and measure the goodness of fit on the same sample.









Left: We examine how well individual subscales explain their corresponding questions. For example how well does the emotional subscale explain variation in the emotion questions?

Right: We examine how well different subscales explain the variation in survey responses.

Key takeaways:

- Removing questions improves goodness of fit particularly for the worst explained subscales.
    - Removing TPFQ Q3/Q8 seems to improve the Emotional subscale
    - Removing TFI Q4 seems to improve the Sense of Control Subscale
- The overall goodness of fit improves most for TPFQ

Conclusions:

The above metrics suggest the surveys from worst to best are: THI < TPFQ < TFI.

<u>Overfitting</u>: The difference between TPFQ and TFI is not always large. Furthermore TFI has several more subscales than TPFQ (8 vs 4). This gives a risk of overstating TFI performance. With additional subscales, TFI has more flexibility to adapt to the data regardless of the true underlying validity. Ideally we would use a measure of goodness of fit which is unaffected by differences in model complexity. Unfortunately classic solutions to this problem (e.g. BIC, cross validation) are inappropriate for our data. Therefore the choice between TPFQ and TFI cannot be made on one metric (e.g. BIC) alone.

BIC is not appropriate because our models use Spearman Correlation and Spearman Correlation does not provide a proper likelihood. Cross validation is only appropriate when we are fitting two models to the same data. In our case each questionnaire is distinct with its own data and therefore cannot be compared.

<u>Removing Questions</u>: Removing select questions seems to improve the goodness of fit of each survey; especially as it relates to the worst performing subscales.

We will later suggest that despite the difficulty that TFI has the best internal validity.

Milestone 1 – Comprehensiveness

In this section we measure how well one questionnaire explains another questionnaire. The most comprehensive questionnaire should explain the others best. In the best case scenario if one questionnaire explains the others perfectly, there is no need to ask the other questionnaires.

Survey to Survey

How well do the subscales from one survey explain all of the subscales from another survey? We fit multiple linear regression and looked at the decrease in variance.

Variation Explained (%)		Predicted Survey THI	Predicted Survey TPFQ	Predicted Survey TFI
Input	THI		50%	47%
	TPFQ	61%		**63%**
	TFI	**65%**	**72%**	

AIC		Predicted Survey THI	Predicted Survey TPFQ	Predicted Survey TFI
Input	THI		2. TPFQ Worse	2. TFI Worse
	TPFQ	2. THI Worse		**1. TFI Better**
	TFI	**1. THI Better**	**1. TPFQ Better**	

Above we see that THI performs worst and TFI/TPFQ perform similarly. TFI is best by a small margin.

Survey to Subscale

Which subscales does a survey predict best? Worst? Although a survey might be the most comprehensive overall, it might not capture specific subscales. This is again done using linear regression, and we looked at the explained variance (%).

Predicting THI — Explained Variance (%)

Predicted Subscale	from TPFQ	from TFI
THI-F_Functional	69%	70%
THI-E_Emotional	51%	61%
THI-C_Catastrophic	36%	36%

TPFQ predicting TFI

Predicted Subscale	Explained Variance (%)
TFI-SL_SLEEP	78%
TFI-C_COGNITIVE	71%
TFI-R_RELAXATION	66%
TFI-Q_QUALITYOFLIFE	62%
TFI-A_AUDITORY	61%
TFI-E_EMOTIONAL	56%
TFI-I_INTRUSIVE	51%
TFI-SC_SENSEOFCONTROL	48%

TFI Predicting TPFQ

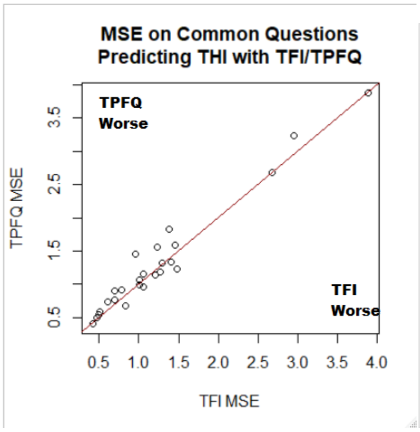
Predicted Subscale	Explained Variance (%)
TPFQ-S_Sleep	78%
TPFQ-C_Concentration	74%
TPFQ-H_Hearing	68%
TPFQ-E_Emotion	67%

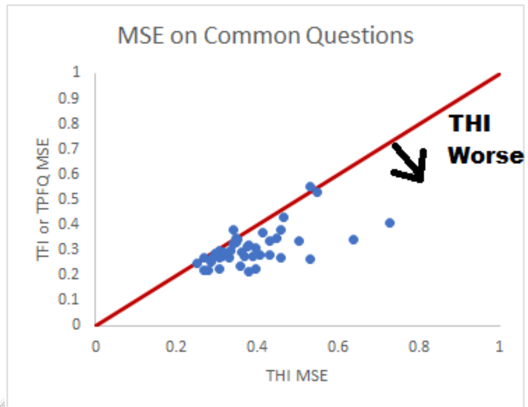
Both TPFQ and TFI provide similar predictions for THI, and both perform poorly on the THI Catastrophic subscale. Since TFI has more subscales than TPFQ, there are some subscales in TFI which TPFQ does not predict well. THI performs poorly with 37%-58% explained variance.

Survey to Question

Which questions does a survey predict best? Worst? Although a survey might be the most comprehensive overall, it might not capture specific questions.

We code all response questions on a 0/2/4 scale with 1-50 coded 2 and 51-100 coded 4. Then for each pair of questionnaires we take all of the subscales from survey A to predict each question from survey B using an ordinal proportional odds model.





Above we see that TFI and TPFQ subscale perform almost identically on THI questions. TFI has a very slight advantage with better MSE in 60% of questions, and better AIC in 52% of questions.

Below we show the questions that TFI and TPFQ predict worst. TFI and TPFQ struggle with the same three questions the most:

ID	Subscale	Questions worst predicted by TFI/TPFQ
THI_Q8	Catastrophic	8. Do you feel as though you cannot escape your tinnitus?
THI_Q19	Catastrophic	19. Do you feel that you have no control over your tinnitus?
THI_Q24	Functional	24. Does your tinnitus get worse when you are under stress?

For additional questions see the appendix (#external)

Conclusion

The above metrics suggest the surveys from worst to best are: THI < TPFQ < TFI. However again the difference between TFI and TPFQ is not large, and can be explained by TFI having more subscales. After controlling for model complexity (e.g. with AIC), the difference between TFI and TPFQ is negligible. One element which neither TFI or TPFQ capture well is the THI Catastrophic subscale.

Milestone 1 – Conclusions

We propose that TFI offers the best internal validity. This is based on the following evidence:

- *Question Strength*: In the question strength section, we identified TFI as the best performer using Cronbach's alpha and EFA. These measures are unaffected by the number of subscales and avoid the 'overfitting' problem above.
- *Subscale Accuracy*: Although the subscale accuracy section is liable to overfit, TFI performs quite strongly. The eight TFI subscales explain 77% of the variance. To match this with the TPFQ not only would we need seven subscales, but those subscales would have to be specially crafted to our observed data and allow for questions to be weighted unequally.
- *Heuristics*: The TFI was the only questionnaire where our custom subscales (EFA) were consistent with the legacy subscales.

In internal validity we assessed how well a questionnaire is described by its subscales. In external validity we measured how well a questionnaire is described by the subscales of another survey. Both approaches suggested the order of the surveys from best to worst is THI < TPFQ < TFI. However the difference between TPFQ and TFI is small and attributable at least in part to the fact that TFI has more subscales (overfitting). It's also possible that if we repeated the experiment that we might instead find a slight advantage for TPFQ by random chance.

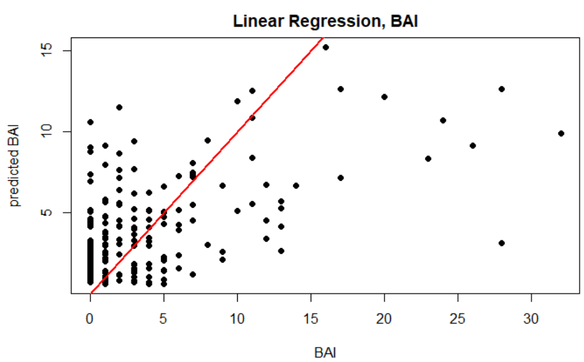
The numbers give a slight advantage to TFI, but not enough to be conclusive. The choice comes down to this. Do you want a slightly more complicated measure (TFI) with more subscales or a slightly simpler/parsimonious measure (TPFQ) with fewer subscales? Both perform similarly.

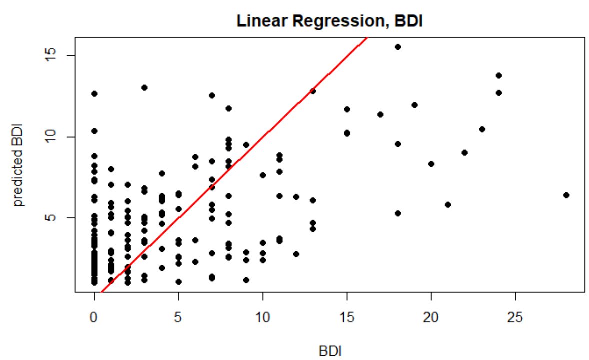
Milestone 2- Emotion

Part 1 Regression

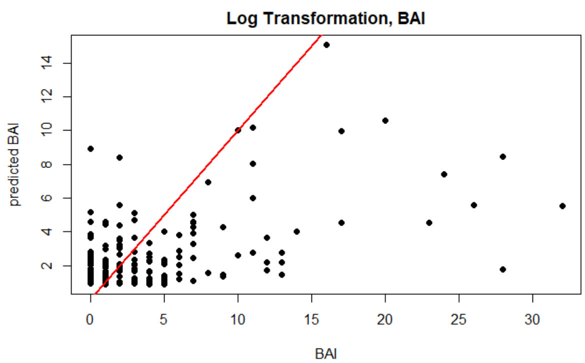
The goal for milestone 2 is to figure out whether it is possible to model the emotional disturbance from tinnitus based on the scores from BDI, BAI and the emotional subscales from the tinnitus annoyance questionnaires. In order to tackle this problem, we have to be clear about our predictors, responses, and problem type. We used BAI and BDI as our response variables. Because all of our data are from questionnaires, they are intrinsically subjective. The most ambiguous part is to find a reasonable and comparably solid value to quantify the emotional disturbance level. Based on our understanding towards the design of each questionnaire, we chose BDI and BAI to be our measure of emotional disturbance. This is because BDI and BAI measure the depression level of each patient due to tunnius. We believe it is reasonable that the depression level somehow describes how this patient's emotion is disturbed by tunnius. While BAI and BDI certainly do not give us a comprehensive view of the true emotional state, they are likely to be the best measures we have in our dataset. The predictors are emotional subscales from TFI, THI, and TPFQ. These subscales are calculated in percentages, therefore this is a regression problem.

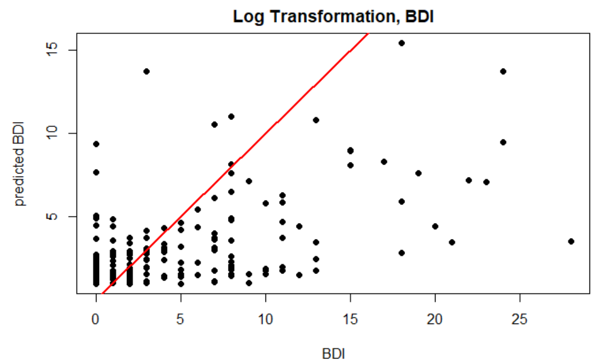
First of all, we extracted the emotional subscales from all three questionnaires, and ran a linear regression on those subscales to predict the total value of BAI and BDI respectively. The predicted values versus the true values are plotted below, and we can see that a simple linear regression does not yield a good result. Specifically, the mean absolute error for predicting BAI is 3.29, while the mean absolute error for predicting BDI is 3.61, which are both pretty large.



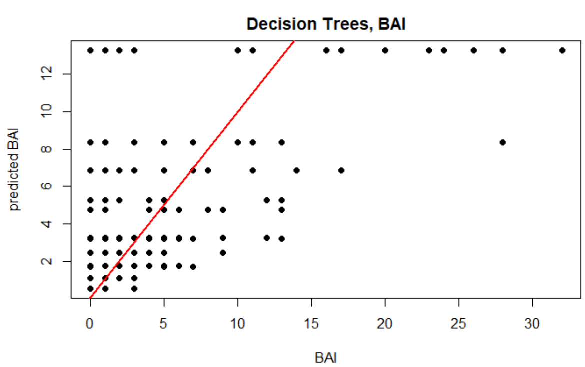


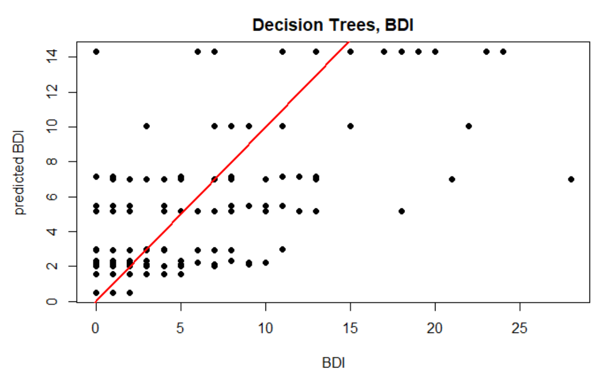
The linear regression models are underfitted due to the fact that the model is not complex enough for our dataset. Since our response variables are highly right-skewed, we tried to transform the BAI and BDI scores into logarithmic scales, in order to neutralize the skewness. The resulting plots still showed underfitting. The mean absolute error for predicting BAI is 3.10, whereas the mean absolute error for predicting BDI is 3.45. Compared to the linear models, we saw a slight improvement, but the errors are still too large to offer us any useful insights.



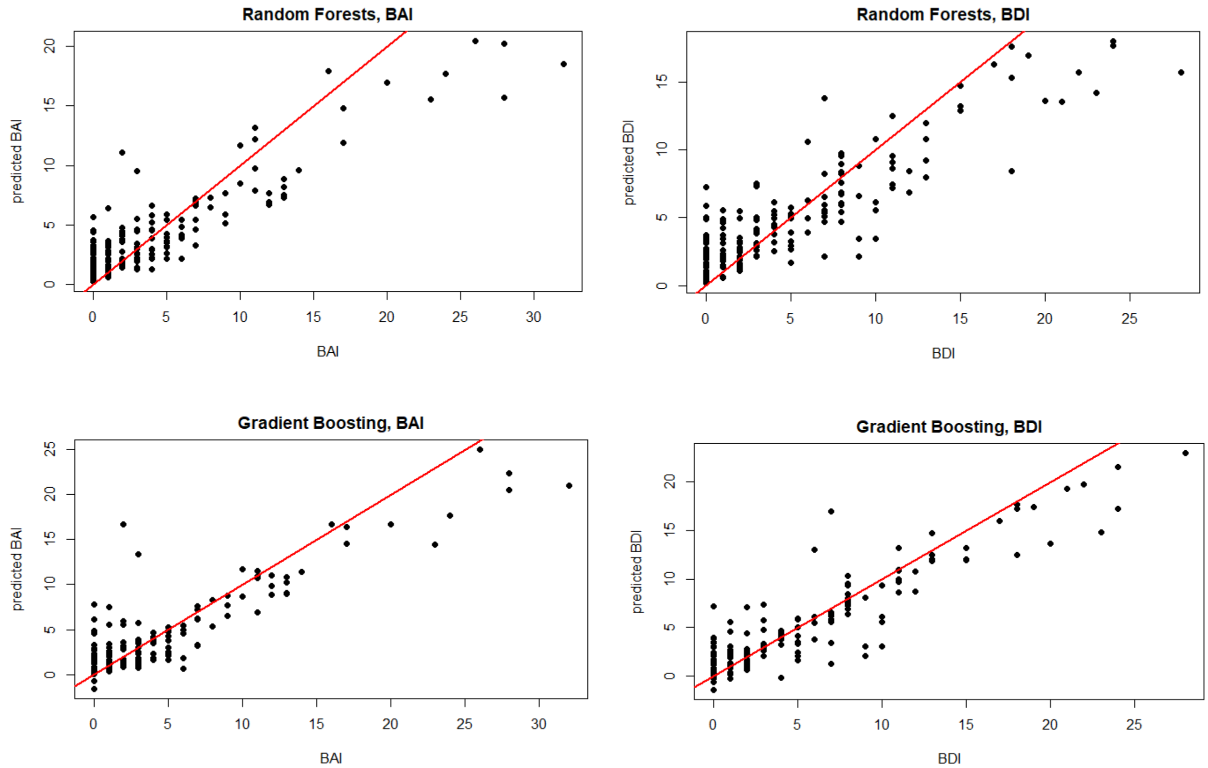


Next, we started to try some non-parametric regression models. The first method we used is the decision tree, which is simply a series of questions. When we get a set of features and values, we use each attribute to answer a question. The answer to each question decides the next question, forming a tree-shaped diagram in the end. In other words, decision trees learn from data to approximate a sine curve with a set of it-then-else decision rules. The deeper the tree, the more complex the decision rules and the fitter the model. The results were still not ideal. The mean absolute error for predicting BAI is 3.00, while the mean absolute error for predicting BDI is 3.10. From the plots below we can see that the decision tree tends to assign discrete values and builds the model in a classification way rather than a regression one.





Then we tried some ensemble methods, which involve using many learners to enhance the performance of any single one of them individually. These methods can be described as techniques that use a group of weak learners together, in order to create a strong, aggregated one. Random Forests are an ensemble of many individual Decision Trees. The mean absolute error for predicting BAI is 1.94, whereas the mean absolute error for predicting BDI is 2.10. We also tried another ensemble method, the stochastic gradient boosting, and it performed even better. The mean absolute error for predicting BAI is 1.69, and the mean absolute error for predicting BDI is 1.58. The plots are shown below.



One of the main drawbacks of ensemble methods such as Random Forests and Stochastic Gradient Boosting is that they are very prone to overfitting: they do well on training data, but are not so flexible for making predictions on unseen samples. Although we have seen improvements in the Random Forests and Stochastic Gradient Boosting methods, they are just training errors. In order for us to conclude that the models can be used, we need to split the data into training set and testing set, then check whether the prediction on the new data can be as good as the training data. From the tables below we can see that the testing mean absolute errors are much higher than the training mean absolute errors. This indicates that Random Forests and Stochastic Gradient Boosting methods are trying too hard to fit on the training data and do not generalize well on new data points, which makes future predictions very hard.

Random Forests	BAI	BDI
Training MAE	1.99	1.64
Testing MAE	**3.47**	**5.20**

Gradient Boosting	BAI	BDI
Training MAE	1.74	1.69

Testing MAE	**2.50**	**4.13**

To sum it up, it is possible to model the emotional disturbance from tinnitus based on BDI, BAI and emotion subscales using techniques like Random Forests and Gradient Boosting, but these models can be highly overfitting and are not useful to predict new observations. In addition, neither BAI score nor BDI score is good enough to be a response variable for this regression task, possibly due to the subjective nature of the self-report questionnaire.

Part 2 Classification

Our goal in this section is to model the emotional disturbance based on the emotional subscales from each questionnaire. The poor fit from our previous regression analysis inspired us to try out classification methods. This is because we only need to identify if the patients are emotionally distrubed or not, and we do not need to consider any other trends or properties embedded in our dataset.

In the classification methods, we started off from a very simple model. To build this model, we used the median in BAI and BDI to be the benchmark to classify patients. Anyone who has a total score more than the benchmark will be classified as 0 (being emotionally distrubed), otherwise, 1 (not being emotionally distrubed). This is because BAI and BDI are extremely skewed to the right. In order to generate a balanced response variable, median is the easiest. We then fitted logistic regressions on classified BAI and BDI separately based on the three emotional subscales. However, this did not give us a good result. To make a more robust response, we combined BAI and BDI together, which means anyone who is being classified as emotionally distruebed in either BAI or BDI will be identified as emotionally disturbed. The combination of BAI and BDI sort of improved our classification sensitivity, but the overall accuracy is still low, which is only a bit above the no information rate (NIR). The results from the simplest model are shown below.

BAI vs 3 Emotional Subscales

Pred BAI	0	1
0	38	48
1	21	80

NIR = 0.54
Accuracy = **0.63**

BDI vs 3 Emotional Subscales

Pred BDI	0	1
0	60	36
1	24	67

NIR = 0.51
Accuracy = **0.68**

BAI & BDI vs 3 Emotional Subscales

Pred BAI & BDI	0	1
0	71	25
1	39	52

NIR = 0.61
Accuracy = **0.66**

In order to improve our model, what we did next was to upgrade our predictors. Based on our results from milestone 1, TFI was basically the best questionnaire, which was not complex and performed the best. Using only three emotional subscales to fit a classification model might not be sufficient in power to explain the classification result from BAI and BDI. So we added any other subscales in TFI into our model and refitted the three models again. The results are printed out below.

BAI vs (3 Emotional Subscales + TFI)

Pred / BAI	0	1
0	47	39
1	26	57

NIR = 0.54
Accuracy = 0.63 ➔ **0.66**

BDI vs (3 Emotional Subscales + TFI)

Pred / BDI	0	1
0	62	34
1	21	70

NIR = 0.51
Accuracy = 0.68 ➔ **0.70**

BAI & BDI vs (3 Emotional Subscales + TFI)

Pred / BAI & BDI	0	1
0	84	30
1	20	53

NIR = 0.61
Accuracy = 0.66 ➔ **0.73**

However, adding extra predictors from TFI did not give good results either. We can see that even if we have a little increase in the prediction accuracy, this increase was too small to offset the expense of adding extra predictors. However, To exhaust every possibility that can improve our prediction accuracy, we would still keep a model with those extra predictors in our future analysis.

Considering we used median as the cut-off and ignored the true nature of BAI and BDI in our previous classification models. We then took a look back at these two questionnaires to see how scientists decided the benchmark. However, the scientific benchmark is way higher than the median, which means there are very few people being classified as "emotionally disturbed", even after we combined the results from BAI and BDI. We call this data unbalanced. In order to fix this problem, we firstly combined the newly classified BAI and BDI together, and then implemented up-sampling methods, SMOTE. Up-sampling methods are helpful in fixing skewed data by generating new data based on existing observations. SOMTE is similar to K-means clustering. It will cluster observations first, then it will randomly generate new observations based on those clusters. After up-sampling by SMOTE, we then fitted a logistic regression model on our balanced predictors. This yielded a good result, which is shown below.

BAI & BDI vs 3 Emotional Subscales

Pred / BAI & BDI	0	1
0	129	40
1	36	138

NIR = 0.50
Accuracy = **0.79**

BAI & BDI vs (3 Emotional Subscales + TFI)

Pred / BAI & BDI	0	1
0	178	2
1	18	163

NIR = 0.50
Accuracy = **0.89**

We can see that after up-sampling, the dataset is balanced and the NIR is 0.5. The accuracy is high, which is around 0.8 for both models. However, we cannot jump into any conclusions now. This is because for our previous analysis in this part, we did fitting and testing on the same dataset, which is potentially to cause an overestimation on the classification accuracy. A cross validation is capable of fixing this problem. However, we cannot do it here. It is because many of the observations are fabricated by SMOTE, there is no meaning if we test the model on the data that is not real. The only way that we can fix this problem is to collect more data, which is potentially to include more observations that are emotionally disturbed into our dataset. Even if there is a problem, classification still has a promising future on modeling the emotional disturbance.

Conclusions

We present the following results from our analysis.

Milestone 1 - Choosing the Best Survey

- We recommend TFI as the primary tool to measure Tinnitus. To further improve validity you could remove questions Q4 and Q22.
- TPFQ is also a good candidate. It's validity is slightly worse, but this is due largely to having fewer subscales. TPFQ validity is improved significantly by removing questions Q3, Q5, and Q8.
- THI has poor fit relative to the other surveys.

Milestone 2 - Influence of Tinnitus onto Emotion

- Neither BAI nor BDI is good enough to be a response variable for regression, possibly due to the subjective nature of the self-report questionnaire.
- Ensemble methods for regression overfit and offer poor predictions.
- Using median as a cutoff, classification on BAI and BDI performs poorly with a simple logistic model.
- After up-sampling, classification fits well. This allows us to use the scientific cutoffs of each questionnaire even though our data is unbalanced.

Appendix

Citations

- (Samejima 1997) https://link.springer.com/chapter/10.1007/978-1-4757-2691-6_5
- (Chalmers 2012) https://www.researchgate.net/publication/228534363_Mirt_A_Multidimensional_Item_Response_Theory_Package_for_the_R_Environment
- Joreskog (1994). KARL G. JORESKOG STRUCTURAL EQUATION MODELINGWITH ORDINAL VARIABLES IMS Lecture Notes - Monograph Series (1994) Volume 24
- Igolkina (2020) Anna A. Igolkina & Georgy Meshcheryakov (2020) semopy: A Python Package for Structural Equation Modeling, Structural Equation Modeling: A Multidisciplinary Journal, 27:6,952-963, DOI: 10.1080/10705511.2019.1704289

Appendix

THI Questions

	SCORE	4	0	2
1.	Does tinnitus impair your concentration?	Yes	No	Sometimes
2.	Does the tinnitus volume make it difficult for you to hear people?	Yes	No	Sometimes
3.	Does tinnitus make you nervous?	Yes	No	Sometimes
4.	Does tinnitus make you confuse?	Yes	No	Sometimes
5.	Are you in despair because of tinnitus?	Yes	No	Sometimes
6.	Do you complain much about your tinnitus?	Yes	No	Sometimes
7.	Do you have difficulties to fall asleep because of your tinnitus?	Yes	No	Sometimes
8.	Do you feel as if you could not escape your tinnitus?	Yes	No	Sometimes
9.	Does your tinnitus impair your social activities (go out for dinner, go to the movies, etc.)?	Yes	No	Sometimes
10.	Are you frustrated because of your tinnitus?	Yes	No	Sometimes
11.	Do you feel as if you had a terrible disease because of your tinnitus?	Yes	No	Sometimes
12.	Does your tinnitus make it difficult for you to enjoy life?	Yes	No	Sometimes
13.	Does your tinnitus interfere with your work and your home chores?	Yes	No	Sometimes
14.	Does your tinnitus make you irritable?	Yes	No	Sometimes
15.	Does your tinnitus impair your reading?	Yes	No	Sometimes
16.	Does your tinnitus make you upset?	Yes	No	Sometimes
17.	Does your tinnitus impair your relationship with family and friends?	Yes	No	Sometimes
18.	Do you have difficulties in turning your attention from the tinnitus to other things?	Yes	No	Sometimes
19.	Do you feel like you had no control over the tinnitus?	Yes	No	Sometimes
20.	Do you feel frequently tired because of your tinnitus?	Yes	No	Sometimes
21.	Do you feel depressed because of your tinnitus?	Yes	No	Sometimes
22.	Does your tinnitus make you stressed?	Yes	No	Sometimes
23.	Do you feel like you could no longer live with your tinnitus?	Yes	No	Sometimes
24.	Does your tinnitus get worse when you are stressed?	Yes	No	Sometimes
25.	Does your tinnitus make you unsure?	Yes	No	Sometimes

TFI Questions

Today's Date ______________ Your Name ______________________

Month / Day / Year — Please Print

Please read each question below carefully. To answer a question, select *ONE* of the numbers that is listed for that question, and draw a *CIRCLE* around it like this: 10% or 1.

I	**Over the PAST WEEK...**

1. What percentage of your time awake were you consciously **AWARE OF** your tinnitus?

Never aware ► 0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100% ◄ *Always aware*

2. How **STRONG** or **LOUD** was your tinnitus?

Not at all strong or loud ► 0 1 2 3 4 5 6 7 8 9 10 ◄ *Extremely strong or loud*

3. What percentage of your time awake were you **ANNOYED** by your tinnitus?

None of the time ► 0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100% ◄ *All of the time*

SC	**Over the PAST WEEK...**

4. Did you feel **IN CONTROL** in regard to your tinnitus?

Very much in control ► 0 1 2 3 4 5 6 7 8 9 10 ◄ *Never in control*

5. How easy was it for you to **COPE** with your tinnitus?

Very easy to cope ► 0 1 2 3 4 5 6 7 8 9 10 ◄ *Impossible to cope*

6. How easy was it for you to **IGNORE** your tinnitus?

Very easy to ignore ► 0 1 2 3 4 5 6 7 8 9 10 ◄ *Impossible to ignore*

C	**Over the PAST WEEK...**

7. Your ability to **CONCENTRATE**?

Did not interfere ► 0 1 2 3 4 5 6 7 8 9 10 ◄ *Completely interfered*

8. Your ability to **THINK CLEARLY?**

Did not interfere ► 0 1 2 3 4 5 6 7 8 9 10 ◄ *Completely interfered*

9. Your ability to **FOCUS ATTENTION** on other things besides your tinnitus?

Did not interfere ► 0 1 2 3 4 5 6 7 8 9 10 ◄ *Completely interfered*

SL	**Over the PAST WEEK...**

10. How often did your tinnitus make it difficult to **FALL ASLEEP or STAY ASLEEP**?

Never had difficulty ► 0 1 2 3 4 5 6 7 8 9 10 ◄ *Always had difficulty*

11. How often did your tinnitus cause you difficulty in getting **AS MUCH SLEEP** as you needed?

Never had difficulty ► 0 1 2 3 4 5 6 7 8 9 10 ◄ *Always had difficulty*

12. How much of the time did your tinnitus keep you from **SLEEPING** as **DEEPLY** or as **PEACEFULLY** as you would have liked?

None of the time ► 0 1 2 3 4 5 6 7 8 9 10 ◄ *All of the time*

 08.15.08

Please read each question below carefully. To answer a question, select *ONE* of the numbers that is listed for that question, and draw a *CIRCLE* around it like this: 10% or 1.

A	**Over the PAST WEEK, how much has your tinnitus interfered with...**	*Did not interfere*										*Completely interfered*
	13. Your ability to **HEAR CLEARLY**?	0	1	2	3	4	5	6	7	8	9	10
	14. Your ability to **UNDERSTAND PEOPLE** who are talking?	0	1	2	3	4	5	6	7	8	9	10
	15. Your ability to **FOLLOW CONVERSATIONS** in a group or at meetings?	0	1	2	3	4	5	6	7	8	9	10

R	**Over the PAST WEEK, how much has your tinnitus interfered with...**	*Did not interfere*										*Completely interfered*
	16. Your **QUIET RESTING ACTIVITIES**?	0	1	2	3	4	5	6	7	8	9	10
	17. Your ability to **RELAX**?	0	1	2	3	4	5	6	7	8	9	10
	18. Your ability to enjoy "**PEACE AND QUIET**"?	0	1	2	3	4	5	6	7	8	9	10

Q	**Over the PAST WEEK, how much has your tinnitus interfered with...**	*Did not interfere*										*Completely interfered*
	19. Your enjoyment of **SOCIAL ACTIVITIES**?	0	1	2	3	4	5	6	7	8	9	10
	20. Your **ENJOYMENT OF LIFE**?	0	1	2	3	4	5	6	7	8	9	10
	21. Your **RELATIONSHIPS** with family, friends and other people?	0	1	2	3	4	5	6	7	8	9	10

22. How often did your tinnitus cause you to have difficulty performing your **WORK OR OTHER TASKS**, such as home maintenance, school work, or caring for children or others?

Never had difficulty ►	0	1	2	3	4	5	6	7	8	9	10	◄ *Always had difficulty*

E Over the PAST WEEK...

23. How **ANXIOUS** or **WORRIED** has your tinnitus made you feel?

Not at all anxious or worried ►	0	1	2	3	4	5	6	7	8	9	10	◄ *Extremely anxious or worried*

24. How **BOTHERED** or **UPSET** have you been because of your tinnitus?

Not at all bothered or upset ►	0	1	2	3	4	5	6	7	8	9	10	◄ *Extremely bothered or upset*

25. How **DEPRESSED** were you because of your tinnitus?

Not at all depressed ►	0	1	2	3	4	5	6	7	8	9	10	◄ *Extremely depressed*

TPFQ Questions

Iowa Tinnitus Primary Function Questionnaire

Name:		Date:	

Please indicate your agreement with each statement on a scale from 0 (completely disagree) to 100 (completely agree).

#	Statement	0-100
1	I have difficulty focusing my attention on some important tasks because of tinnitus.	
2	I lie awake at night because of my tinnitus.	
3	I just wish my tinnitus would go away. It is so frustrating.	
4	I have difficulty getting to sleep at night because of my tinnitus.	
5	When there are lots of things happening at once, my tinnitus interferes with my ability to attend to the most important thing.	
6	My tinnitus masks some speech sounds.	
7	My inability to think about something undisturbed is one of the worst effects of my tinnitus.	
8	My tinnitus is annoying.	
9	One of the worst things about my tinnitus is its effect on my speech understanding, over and above any effect of my hearing loss.	
10	My tinnitus, not my hearing loss, interferes with my appreciation of music and songs.	
11	I am tired during the day because my tinnitus has disrupted my sleep.	
12	In addition to my hearing loss, my tinnitus interferes with my understanding of speech.	
13	I am depressed because of my tinnitus.	
14	When I wake up in the night, my tinnitus makes it difficult to get back to sleep.	
15	My emotional peace is one of the worst effects of my tinnitus.	
16	I have trouble concentrating while I am reading in a quiet room because of tinnitus.	
17	The difficulty I have sleeping is one of the worst effects of my tinnitus.	
18	I am anxious because of my tinnitus.	
19	The effects of tinnitus on my hearing are worse than the effects of my hearing loss.	
20	I feel like my tinnitus makes it difficult for me to concentrate on some tasks.	

Questions with subscales:

TFI

ID	Question	Subscale
Q1	1. What percentage of your time awake were you consciously AWARE OF your tinnitus?	INTRUSIVE
Q2	2. How STRONG or LOUD was your tinnitus?	INTRUSIVE
Q3	3. What percentage of your time awake were you ANNOYED by your tinnitus?	INTRUSIVE
Q4	4. Did you feel IN CONTROL in regard to your tinnitus?	SENSE OF CONTROL
Q5	5. How easy was it for you to COPE with your tinnitus?	SENSE OF CONTROL
Q6	6. How easy was it for you to IGNORE your tinnitus?	SENSE OF CONTROL
Q7	7. Your ability to CONCENTRATE?	COGNITIVE
Q8	8. Your ability to THINK CLEARLY?	COGNITIVE
Q9	9. Your ability to FOCUS ATTENTION on other things besides your tinnitus?	COGNITIVE
Q10	10. How often did your tinnitus make it difficult to FALL ASLEEP or STAY ASLEEP?	SLEEP
Q11	11. How often did your tinnitus cause you difficulty in getting AS MUCH SLEEP as you needed?	SLEEP
Q12	12. How much of the time did your tinnitus keep you from SLEEPING as DEEPLY or as PEACEFULLY as you would have liked?	SLEEP
Q13	13. Your ability to HEAR CLEARLY?	AUDITORY
Q14	14. Your ability to UNDERSTAND PEOPLE who are talking?	AUDITORY
Q15	15. Your ability to FOLLOW CONVERSATIONS in a group or at meetings?	AUDITORY
Q16	16. Your QUIET RESTING ACTIVITIES?	RELAXATION
Q17	17. Your ability to RELAX?	RELAXATION
Q18	18. Your ability to enjoy "PEACE AND QUIET"?	RELAXATION
Q19	19. Your enjoyment of SOCIAL ACTIVITIES?	QUALITY OF LIFE
Q20	20. Your ENJOYMENT OF LIFE?	QUALITY OF LIFE
Q21	21. Your RELATIONSHIPS with family, friends and other people?	QUALITY OF LIFE
Q22	22. How often did your tinnitus cause you to have difficulty performing your WORK OR OTHER TASKS, such as home maintenance, school work, or caring for children or others?	QUALITY OF LIFE
Q23	23. How ANXIOUS or WORRIED has your tinnitus made you feel?	EMOTIONAL
Q24	24. How BOTHERED or UPSET have you been because of your tinnitus?	EMOTIONAL
Q25	25. How DEPRESSED were you because of your tinnitus?	EMOTIONAL

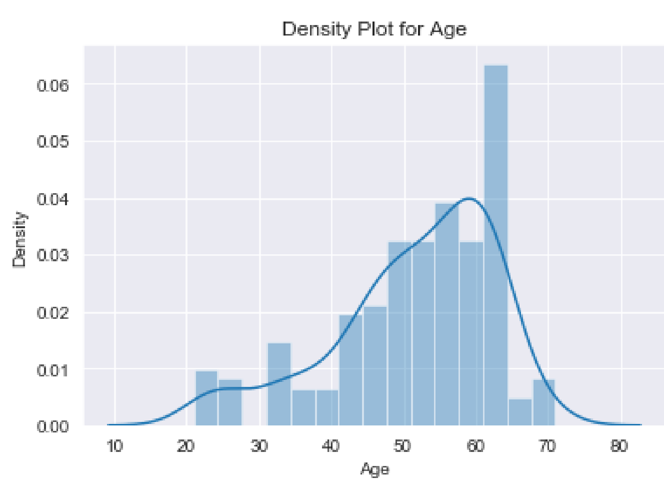
THI:

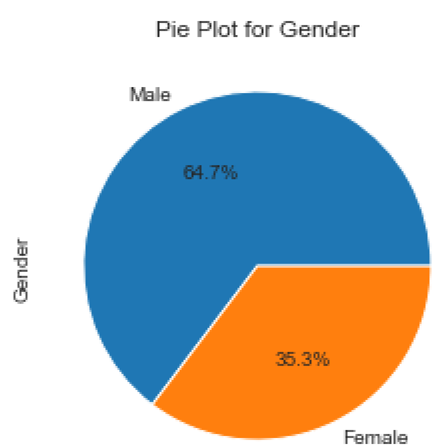
ID	Question	Subscale
Q1	1. Because of your tinnitus, is it difficult for you to concentrate?	Functional
Q2	2. Does the loudness of your tinnitus make it difficult for you to hear people?	Functional
Q4	4. Does your tinnitus make you feel confused?	Functional
Q7	7. Because of your tinnitus, do you have trouble falling to sleep at night?	Functional
Q9	9. Does your tinnitus interfere with your ability to enjoy your social activities (such as going out to dinner, to the movies)?	Functional
Q12	12. Does your tinnitus make it difficult for you to enjoy life?	Functional
Q13	13. Does your tinnitus interfere with your job or household responsibilities?	Functional
Q14	14. Because of your tinnitus, do you find that you are often irritable?	Functional
Q15	15. Because of your tinnitus, is it difficult for you to read?	Functional
Q18	18. Do you find it difficult to focus your attention away from your tinnitus and on other things?	Functional
Q20	20. Because of your tinnitus, do you often feel tired?	Functional
Q24	24. Does your tinnitus get worse when you are under stress?	Functional
Q3	3. Does your tinnitus make you angry?	Emotional
Q6	6. Do you complain a great deal about your tinnitus?	Emotional
Q10	10. Because of your tinnitus, do you feel frustrated?	Emotional
Q16	16. Does your tinnitus make you upset?	Emotional
Q17	17. Do you feel that your tinnitus problem has placed stress on your relationships with members of your family and friends?	Emotional
Q21	21. Because of your tinnitus, do you feel depressed?	Emotional
Q22	22. Does your tinnitus make you feel anxious?	Emotional
Q25	25. Does your tinnitus make you feel insecure?	Emotional
Q5	5. Because of your tinnitus, do you feel desperate?	Catastrophic
Q8	8. Do you feel as though you cannot escape your tinnitus?	Catastrophic
Q11	11. Because of your tinnitus, do you feel that you have a terrible disease?	Catastrophic
Q19	19. Do you feel that you have no control over your tinnitus?	Catastrophic

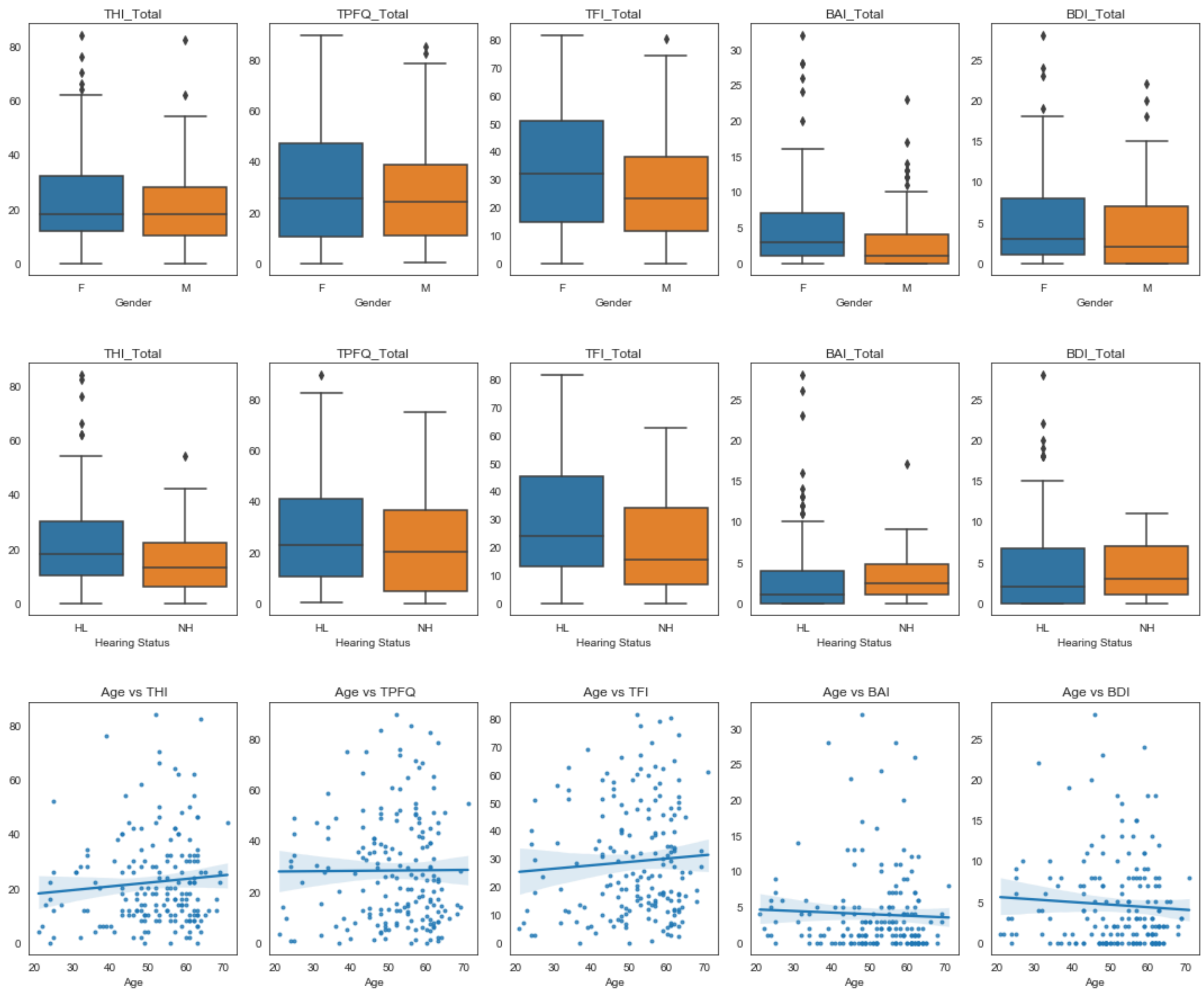
TPFQ

ID	Question	Subscale
Q1	1 I have difficulty focusing my attention on some important tasks because of tinnitus.	Concentration
Q5	5 When there are lots of things happening at once, my tinnitus interferes with my ability to attend to the most important thing.	Concentration
Q7	7 My inability to think about something undisturbed is one of the worst effects of my tinnitus.	Concentration
Q16	16 I have trouble concentrating while I am reading in a quiet room because of tinnitus.	Concentration
Q20	20 I feel like my tinnitus makes it difficult for me to concentrate on some tasks.	Concentration
Q2	2 I lie awake at night because of my tinnitus.	Sleep
Q4	4 I have difficulty getting to sleep at night because of my tinnitus.	Sleep
Q11	11 I am tired during the day because my tinnitus has disrupted my sleep.	Sleep
Q14	14 When I wake up in the night, my tinnitus makes it difficult to get back to sleep.	Sleep
Q17	17 The difficulty I have sleeping is one of the worst effects of my tinnitus.	Sleep
Q3	3 I just wish my tinnitus would go away. It is so frustrating.	Emotion
Q8	8 My tinnitus is annoying.	Emotion
Q13	13 I am depressed because of my tinnitus.	Emotion
Q15	15 My emotional peace is one of the worst effects of my tinnitus.	Emotion
Q18	18 I am anxious because of my tinnitus.	Emotion
Q6	6 My tinnitus masks some speech sounds.	Hearing
Q9	9 One of the worst things about my tinnitus is its effect on my loss. speech understanding, over and above any effect of my hearing	Hearing
Q10	10 My tinnitus, not my hearing loss, interferes with my appreciation of music and songs.	Hearing
Q12	12 In addition to my hearing loss, my tinnitus interferes with my understanding of speech.	Hearing
Q19	19 The effects of tinnitus on my hearing are worse than the effects of my hearing loss.	Hearing

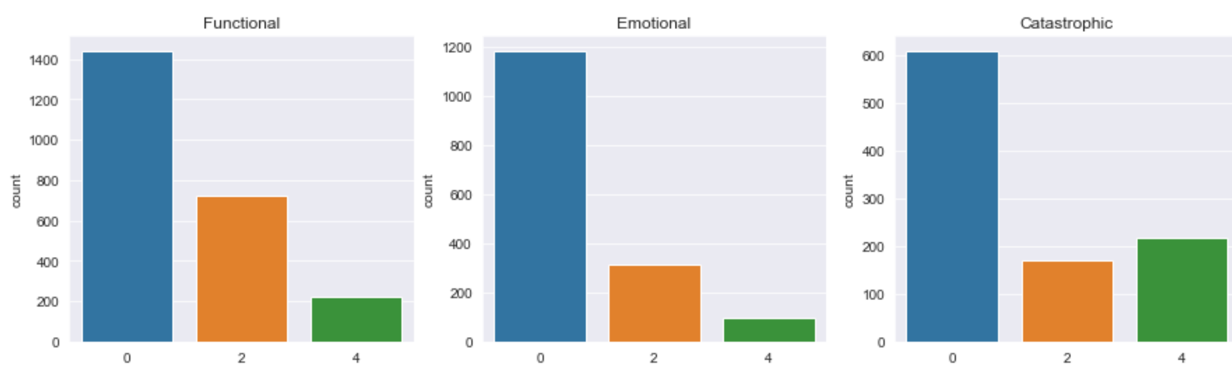
EDA



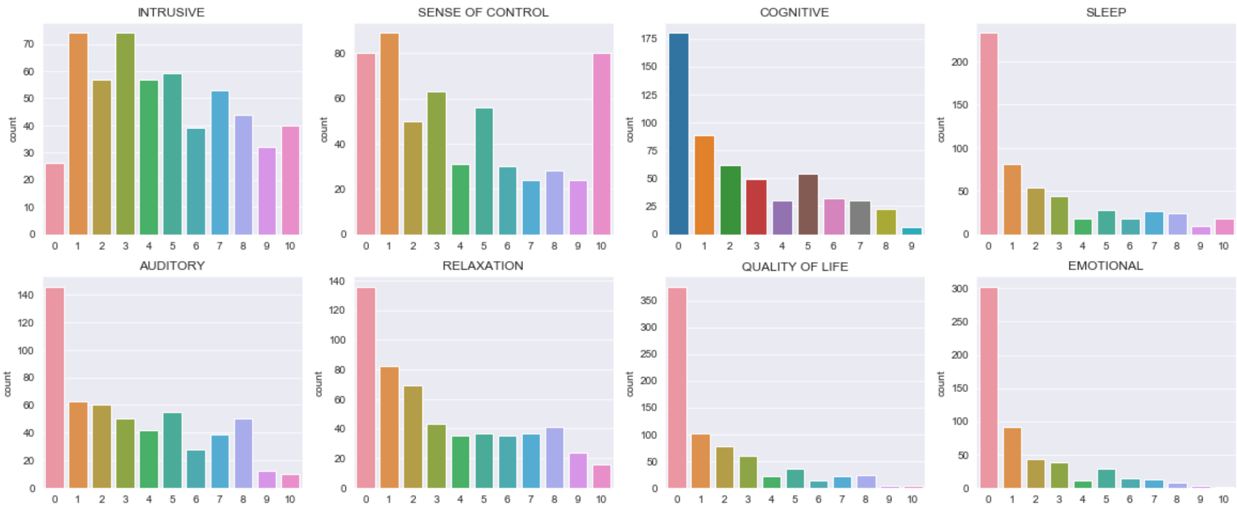



THI_Total
TPFQ_Total
TFI_Total
BAI_Total
BDI_Total
Gender
F
M
Hearing Status
HL
NH
Age vs THI
Age vs TPFQ
Age vs TFI
Age vs BAI
Age vs BDI
Age

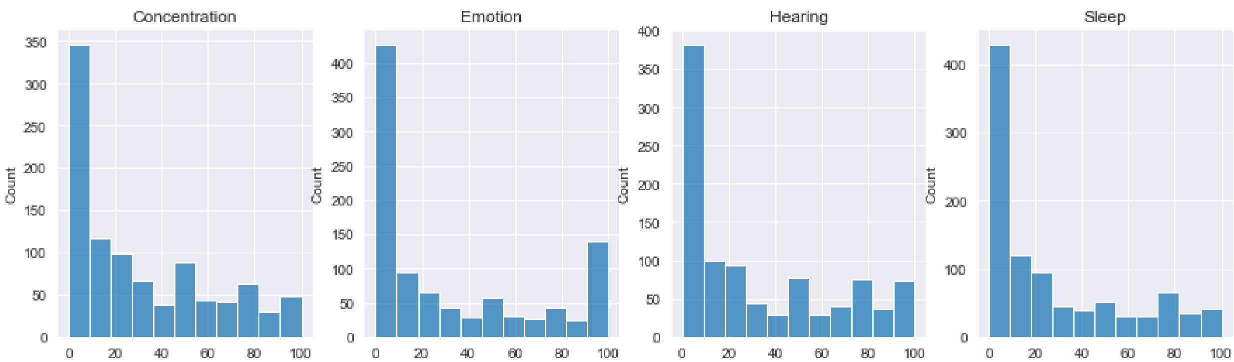
THI


Functional
Emotional
Catastrophic
count

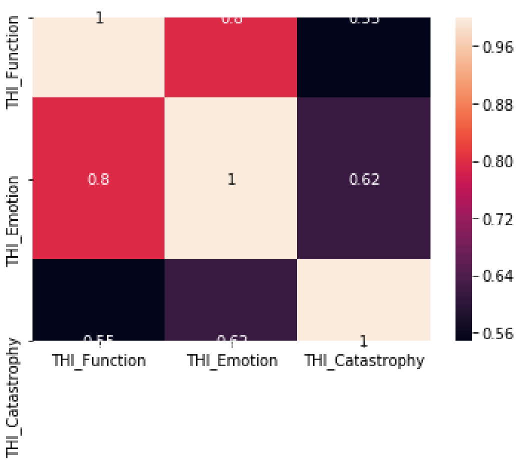
TFI

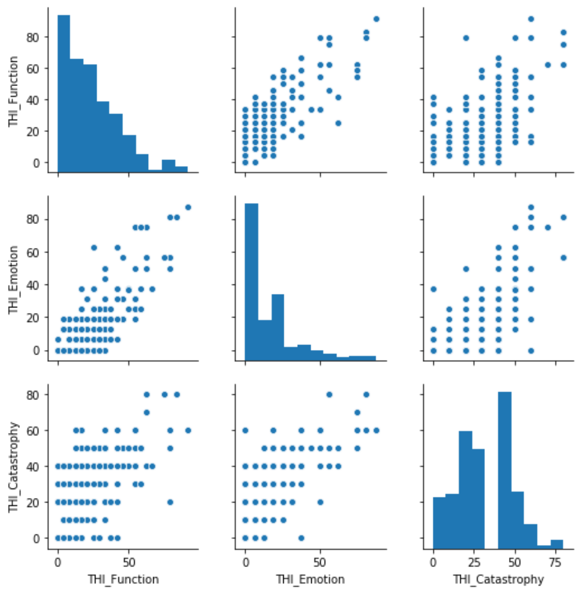


TPFQ

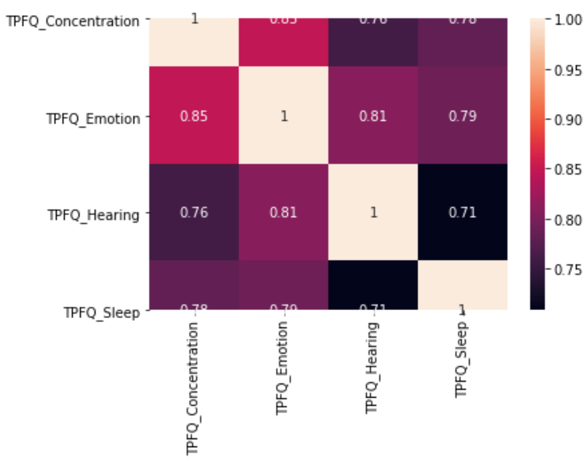


Within Questionnaire Subscale Correlations

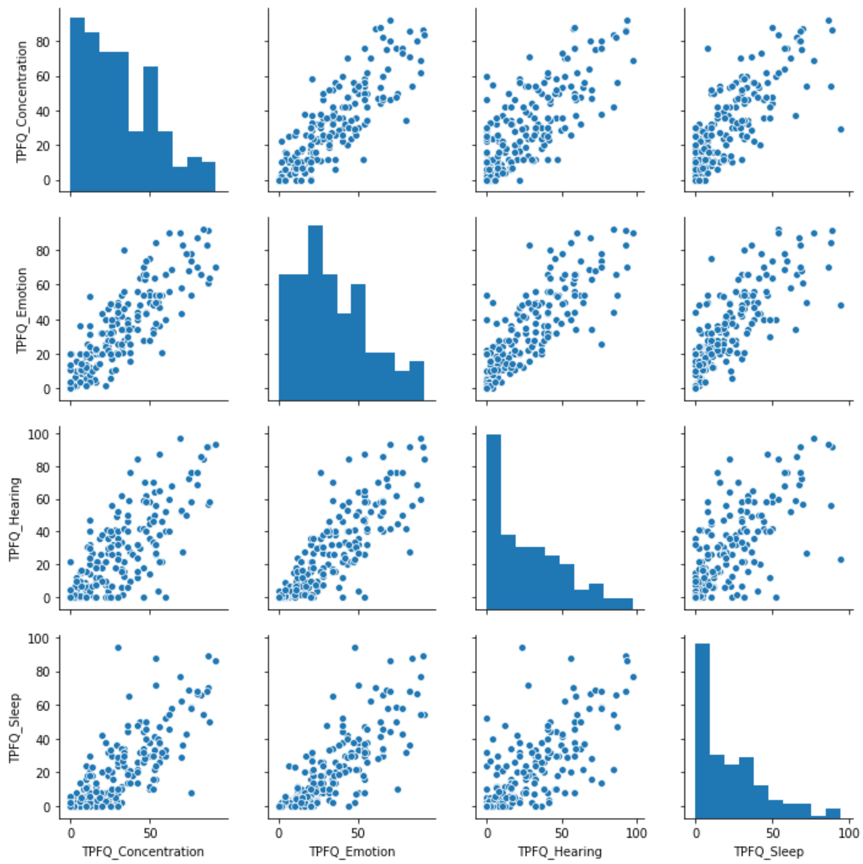


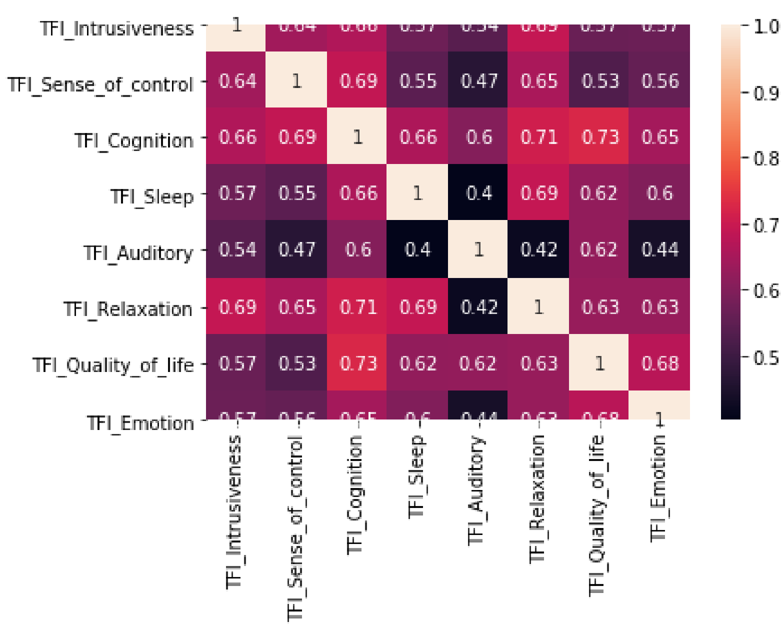


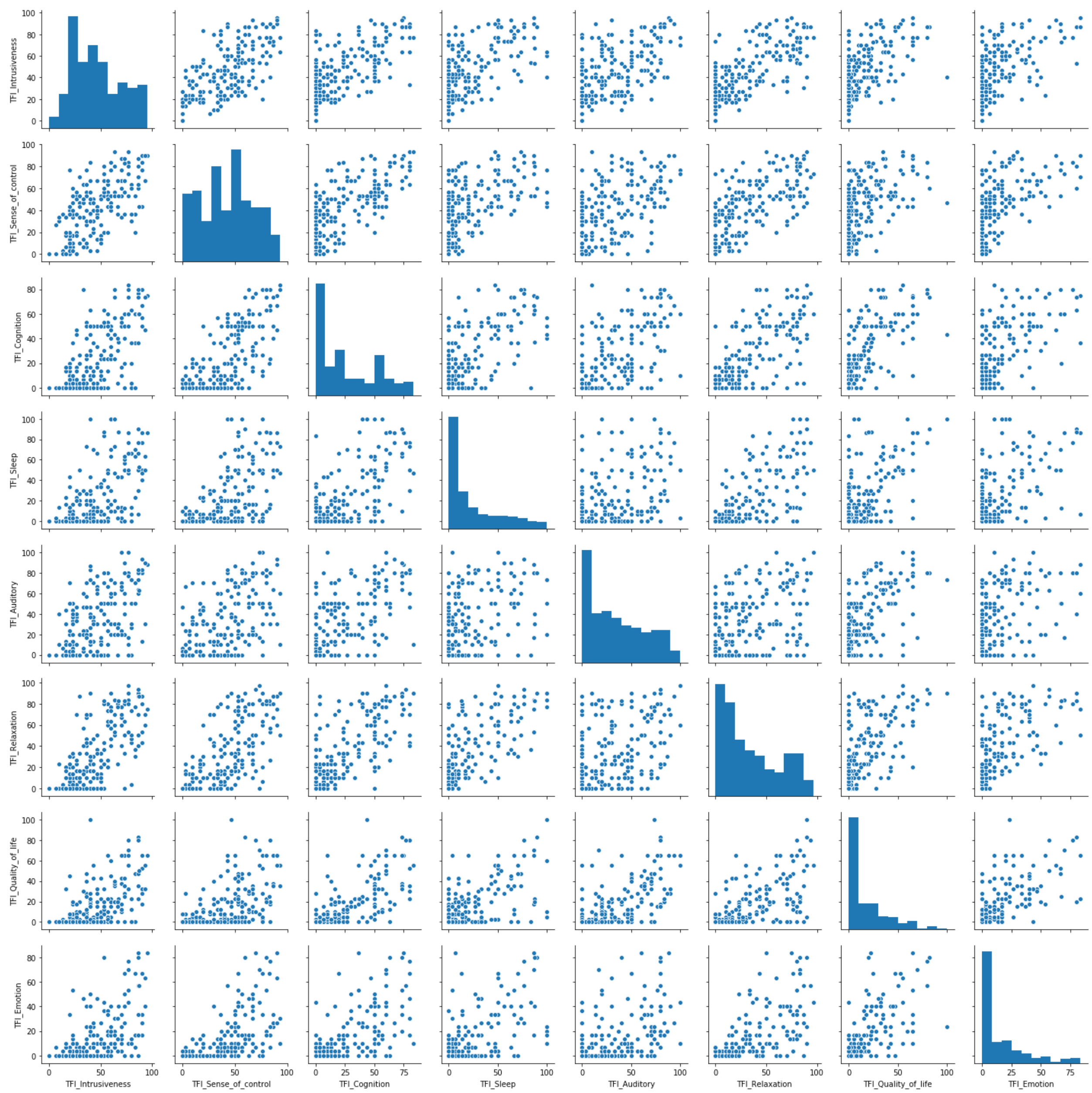
THI_Function
THI_Emotion
THI_Catastrophy



TPFQ_Concentration
TPFQ_Emotion
TPFQ_Hearing
TPFQ_Sleep
0.85
0.81
0.79
0.76
0.71


TPFQ_Concentration
TPFQ_Emotion
TPFQ_Hearing
TPFQ_Sleep
TPFQ_Concentration
TPFQ_Emotion
TPFQ_Hearing
TPFQ_Sleep


TFI_Intrusiveness
TFI_Sense_of_control
TFI_Cognition
TFI_Sleep
TFI_Auditory
TFI_Relaxation
TFI_Quality_of_life
TFI_Emotion
1
0.64 1 0.69 0.55 0.47 0.65 0.53 0.56
0.66 0.69 1 0.66 0.6 0.71 0.73 0.65
0.57 0.55 0.66 1 0.4 0.69 0.62 0.6
0.54 0.47 0.6 0.4 1 0.42 0.62 0.44
0.69 0.65 0.71 0.69 0.42 1 0.63 0.63
0.57 0.53 0.73 0.62 0.62 0.63 1 0.68
1
1.0
0.9
0.8
0.7
0.6
0.5
TFI_Intrusiveness
TFI_Sense_of_control
TFI_Cognition
TFI_Sleep
TFI_Auditory
TFI_Relaxation
TFI_Quality_of_life
TFI_Emotion



Quality Measures (#XMeasures)

TK fill in for Milestone 1 Internal validity

External Validity (#external)

Below we fit individual questions of one survey with the subscales of another survey using a proportional odds model. In each case the responses are broken out into three categories corresponding to: 0, 1-50, 51-100.

Input Subscales	Output	Output Questions	MSE	Baseline MSE Intercept Only	AIC
TFI	THI	THI_Q1	1.30	1.90	258.1
TPFQ	THI	THI_Q1	1.32	1.90	266.1
TFI	THI	THI_Q2	1.45	2.39	278.5
TPFQ	THI	THI_Q2	1.59	2.39	288.3
TFI	THI	THI_Q3	1.39	1.67	233.1
TPFQ	THI	THI_Q3	1.83	1.67	239.3
TFI	THI	THI_Q4	0.69	0.85	150.3
TPFQ	THI	THI_Q4	0.76	0.85	146.9
TFI	THI	THI_Q5	0.42	0.42	120.1
TPFQ	THI	THI_Q5	0.40	0.42	98.3
TFI	THI	THI_Q6	1.23	1.54	231.0
TPFQ	THI	THI_Q6	1.56	1.54	245.0
TFI	THI	THI_Q7	1.27	2.12	262.1
TPFQ	THI	THI_Q7	1.18	2.12	237.2
TFI	THI	THI_Q8	2.68	2.77	346.5
TPFQ	THI	THI_Q8	2.68	2.77	341.6

TFI	THI	THI_Q9	1.41	1.80	239.2
TPFQ	THI	THI_Q9	1.34	1.80	252.2
TFI	THI	THI_Q10	1.47	1.98	276.9
TPFQ	THI	THI_Q10	1.23	1.98	258.1
TFI	THI	THI_Q11	0.47	0.51	102.3
TPFQ	THI	THI_Q11	0.49	0.51	96.2
TFI	THI	THI_Q12	0.83	1.27	165.7
TPFQ	THI	THI_Q12	0.67	1.27	173.1
TFI	THI	THI_Q13	0.69	1.31	165.1
TPFQ	THI	THI_Q13	0.89	1.31	184.2
TFI	THI	THI_Q14	1.01	1.36	204.2
TPFQ	THI	THI_Q14	1.07	1.36	203.8
TFI	THI	THI_Q15	1.21	1.49	252.6
TPFQ	THI	THI_Q15	1.14	1.49	236.2
TFI	THI	THI_Q16	0.96	1.45	192.9
TPFQ	THI	THI_Q16	1.45	1.45	214.5
TFI	THI	THI_Q17	1.05	1.49	186.9
TPFQ	THI	THI_Q17	1.16	1.49	196.6

TFI	THI	THI_Q18	1.01	1.27	219.6
TPFQ	THI	THI_Q18	0.98	1.27	201.2
TFI	THI	THI_Q19	2.95	2.39	326.4
TPFQ	THI	THI_Q19	3.24	2.39	332.7
TFI	THI	THI_Q20	1.05	1.54	210.5
TPFQ	THI	THI_Q20	0.96	1.54	194.6
TFI	THI	THI_Q21	0.60	0.83	109.8
TPFQ	THI	THI_Q21	0.74	0.83	114.4
TFI	THI	THI_Q22	0.78	1.32	179.3
TPFQ	THI	THI_Q22	0.92	1.32	173.8
TFI	THI	THI_Q23	0.49	0.60	113.7
TPFQ	THI	THI_Q23	0.54	0.60	115.5
TFI	THI	THI_Q24	3.89	2.47	352.6
TPFQ	THI	THI_Q24	3.89	2.47	344.2
TFI	THI	THI_Q25	0.51	0.60	107.0
TPFQ	THI	THI_Q25	0.58	0.60	118.4
THI	TFI	TFI_Q1....	0.35	0.50	264.5
TPFQ	TFI	TFI_Q1....	0.34	0.50	256.1

THI	TFI	TFI_Q2			
TPFQ	TFI	TFI_Q2			
THI	TFI	TFI_Q3....	0.25	0.32	220.4062
TPFQ	TFI	TFI_Q3....	0.25	0.32	206.4081
THI	TFI	TFI_Q4	0.53	0.67	314.7928
TPFQ	TFI	TFI_Q4	0.55	0.67	331.1093
THI	TFI	TFI_Q5	0.31	0.35	255.7714
TPFQ	TFI	TFI_Q5	0.27	0.35	248.377
THI	TFI	TFI_Q6	0.34	0.41	278.5003
TPFQ	TFI	TFI_Q6	0.30	0.41	255.21
THI	TFI	TFI_Q7	0.31	0.46	257.6416
TPFQ	TFI	TFI_Q7	0.27	0.46	234.3514
THI	TFI	TFI_Q8	0.32	0.55	253.8266
TPFQ	TFI	TFI_Q8	0.29	0.55	239.1421
THI	TFI	TFI_Q9	0.33	0.47	263.564
TPFQ	TFI	TFI_Q9	0.27	0.47	229.067
THI	TFI	TFI_Q10	0.46	0.58	328.9656
TPFQ	TFI	TFI_Q10	0.27	0.58	222.3961

THI	TFI	TFI_Q11	0.38	0.59	278.5193
TPFQ	TFI	TFI_Q11	0.21	0.59	191.8528
THI	TFI	TFI_Q12	0.41	0.60	293.809
TPFQ	TFI	TFI_Q12	0.28	0.60	231.8441
THI	TFI	TFI_Q13	0.40	0.48	307.0644
TPFQ	TFI	TFI_Q13	0.31	0.48	257.9188
THI	TFI	TFI_Q14	0.45	0.53	312.0856
TPFQ	TFI	TFI_Q14	0.35	0.53	273.4687
THI	TFI	TFI_Q15	0.43	0.54	317.2032
TPFQ	TFI	TFI_Q15	0.34	0.54	273.1275
THI	TFI	TFI_Q16	0.38	0.53	302.0177
TPFQ	TFI	TFI_Q16	0.32	0.53	261.7775
THI	TFI	TFI_Q17	0.37	0.54	284.5789
TPFQ	TFI	TFI_Q17	0.31	0.54	255.5069
THI	TFI	TFI_Q18	0.34	0.50	292.4317
TPFQ	TFI	TFI_Q18	0.32	0.50	259.8541
THI	TFI	TFI_Q19	0.36	0.62	262.5271
TPFQ	TFI	TFI_Q19	0.29	0.62	245.3333

THI	TFI	TFI_Q20	0.31	0.59	240.9554
TPFQ	TFI	TFI_Q20	0.30	0.59	230.3639
THI	TFI	TFI_Q21	0.35	0.61	265.7393
TPFQ	TFI	TFI_Q21	0.33	0.61	253.1778
THI	TFI	TFI_Q22	0.30	0.59	228.2657
TPFQ	TFI	TFI_Q22	0.28	0.59	238.2014
THI	TFI	TFI_Q23	0.27	0.61	220.7706
TPFQ	TFI	TFI_Q23	0.27	0.61	224.8454
THI	TFI	TFI_Q24	0.35	0.47	257.1141
TPFQ	TFI	TFI_Q24	0.35	0.47	262.9568
THI	TFI	TFI_Q25	0.28	0.44	202.3497
TPFQ	TFI	TFI_Q25	0.25	0.44	190.6957
THI	TPFQ	TPFQ_Q1	0.28	0.47	259.5626
TFI	TPFQ	TPFQ_Q1	0.22	0.47	209.8074
THI	TPFQ	TPFQ_Q2	0.40	0.51	309.5499
TFI	TPFQ	TPFQ_Q2	0.22	0.51	226.4124
THI	TPFQ	TPFQ_Q3	0.30	0.55	262.2938
TFI	TPFQ	TPFQ_Q3			

THI	TPFQ	TPFQ_Q4	0.39	0.54	318.4813
TFI	TPFQ	TPFQ_Q4	0.27	0.54	241.5564
THI	TPFQ	TPFQ_Q5	0.37	0.56	274.062
TFI	TPFQ	TPFQ_Q5	0.27	0.56	239.2041
THI	TPFQ	TPFQ_Q6	0.73	0.68	337.5809
TFI	TPFQ	TPFQ_Q6	0.41	0.68	280.1652
THI	TPFQ	TPFQ_Q7	0.46	0.57	315.7199
TFI	TPFQ	TPFQ_Q7	0.43	0.57	310.6259
THI	TPFQ	TPFQ_Q8			
TFI	TPFQ	TPFQ_Q8			
THI	TPFQ	TPFQ_Q9	0.50	0.63	317.6509
TFI	TPFQ	TPFQ_Q9	0.34	0.63	256.437
THI	TPFQ	TPFQ_Q10	0.46	0.62	311.825
TFI	TPFQ	TPFQ_Q10	0.38	0.62	282.5465
THI	TPFQ	TPFQ_Q11	0.36	0.62	269.0303
TFI	TPFQ	TPFQ_Q11	0.23	0.62	219.9973
THI	TPFQ	TPFQ_Q12	0.53	0.61	325.3018
TFI	TPFQ	TPFQ_Q12	0.26	0.61	245.1357

THI	TPFQ	TPFQ_Q13	0.27	0.34	198.8964
TFI	TPFQ	TPFQ_Q13	0.22	0.34	195.9135
THI	TPFQ	TPFQ_Q14	0.43	0.59	300.8011
TFI	TPFQ	TPFQ_Q14	0.28	0.59	261.9814
THI	TPFQ	TPFQ_Q15	0.34	0.71	258.6999
TFI	TPFQ	TPFQ_Q15	0.38	0.71	275.0589
THI	TPFQ	TPFQ_Q16	0.41	0.55	310.6283
TFI	TPFQ	TPFQ_Q16	0.37	0.55	287.3266
THI	TPFQ	TPFQ_Q17	0.64	0.67	322.5982
TFI	TPFQ	TPFQ_Q17	0.34	0.67	260.9343
THI	TPFQ	TPFQ_Q18	0.31	0.50	204.4669
TFI	TPFQ	TPFQ_Q18	0.22	0.50	208.6798
THI	TPFQ	TPFQ_Q19	0.55	0.66	320.5109
TFI	TPFQ	TPFQ_Q19	0.53	0.66	314.2372
THI	TPFQ	TPFQ_Q20	0.29	0.48	261.0669
TFI	TPFQ	TPFQ_Q20	0.27	0.48	238.3988

Emotion Scores Across Questionnaires

BAI Q19 Faint / lightheaded

```
                                    Estimate Std. Error z value Pr(>|z|)
(Intercept)                         -8.97215    3.76632  -2.382   0.0172 *
xdf[nafilter, ]THI_E_Emotional       0.52070    0.24412   2.133   0.0329 *
xdf[nafilter, ]THI_C_Catastrophic -0.63202    0.42736  -1.479   0.1392
xdf[nafilter, ]TFI_E_EMOTIONAL     -0.11114    0.18073  -0.615   0.5386
xdf[nafilter, ]TPFQ_E_Emotion       0.01309    0.01363   0.960   0.3369
---
               Estimate Std. Error z value Pr(>|z|)
(Intercept)     -7.54151    2.09383  -3.602 0.000316 ***
THI_E_Emotional  0.27138    0.09608   2.825 0.004734 **
```

BAI Q18: Appetite

```
                                     Estimate Std. Error z value Pr(>|z|)
(Intercept)                         -2.303166   0.499689  -4.609 4.04e-06 ***
xdf[nafilter, ]THI_E_Emotional       0.043089   0.047045   0.916    0.360
xdf[nafilter, ]THI_C_Catastrophic  0.096332   0.079481   1.212    0.226
xdf[nafilter, ]TFI_E_EMOTIONAL      0.077719   0.050793   1.530    0.126
xdf[nafilter, ]TPFQ_E_Emotion      -0.001882   0.003271  -0.575    0.565
---
               Estimate Std. Error z value Pr(>|z|)
(Intercept)     -1.90434    0.26186  -7.272 3.53e-13 ***
TFI_E_EMOTIONAL  0.10980    0.03024   3.631 0.000282 ***
```

TFI

BAI Q19 Faint / lightheaded

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-250.338	78048.379	-0.003	0.997
xdf[nafilter,]TFI_I_INTRUSIVE	-8.852	7199.448	-0.001	0.999
xdf[nafilter,]TFI_SC_SENSEOFCONTROL	2.110	9640.982	0.000	1.000
xdf[nafilter,]TFI_C_COGNITIVE	2.947	12093.300	0.000	1.000
xdf[nafilter,]TFI_SL_SLEEP	-3.042	5948.178	-0.001	1.000
xdf[nafilter,]TFI_A_AUDITORY	6.602	5396.141	0.001	0.999
xdf[nafilter,]TFI_R_RELAXATION	3.938	3297.256	0.001	0.999
xdf[nafilter,]TFI_Q_QUALITYOFLIFE	3.230	2988.402	0.001	0.999
xdf[nafilter,]TFI_E_EMOTIONAL	2.782	9761.085	0.000	1.000

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-9.5469	3.2983	-2.894	0.0038	**
TFI_Q_QUALITYOFLIFE	0.2775	0.1143	2.428	0.0152	*

BAI Q11 Feeling of choking

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-57.816	6964.487	-0.008	0.993
xdf[nafilter,]TFI_I_INTRUSIVE	-1.757	564.399	-0.003	0.998
xdf[nafilter,]TFI_SC_SENSEOFCONTROL	-82.061	6062.447	-0.014	0.989
xdf[nafilter,]TFI_C_COGNITIVE	6.232	1582.589	0.004	0.997
xdf[nafilter,]TFI_SL_SLEEP	16.933	1304.401	0.013	0.990
xdf[nafilter,]TFI_A_AUDITORY	-1.405	1299.535	-0.001	0.999
xdf[nafilter,]TFI_R_RELAXATION	17.313	1357.075	0.013	0.990
xdf[nafilter,]TFI_Q_QUALITYOFLIFE	1.986	795.316	0.002	0.998
xdf[nafilter,]TFI_E_EMOTIONAL	20.281	1577.301	0.013	0.990

EFA

For additional information on Exploratory Factor analysis see the Excel file provided here on Box. This will be available there through 8/1/21.

Code

The code used to generate the report can be found here on Box and will be available there through 8/1/21.

Authors

	Primary Author	**Editor**
Background	Zhehui	Shuyu
EDA	Zhehui	Shuyu
Methods	Mix	Yumeng
Milestone 1		
EFA	Yumeng	Jesse
CFA	Jesse	Yumeng
Milestone 2		
Regression	Shuyu	Shiwei
Classification	Shiwei	Zhehui

Final Conclusion	Mix	Mix