EXPLORATORY FACTOR ANALYSIS: MODEL SELECTION AND IDENTIFYING UNDERLYING SYMPTOMS

By

Matthew K. Cole

A thesis submitted to Johns Hopkins University in conformity with the requirements for the degree of Master of Science.

Baltimore, Maryland

September, 2017

Abstract:

Exploratory factor analysis (EFA) is a common yet powerful tool to better understand the theoretical structure of a set of variables. A core problem of conducting an EFA is determining the number of factors (*m*) to extract and examine. In this thesis, we examined the performance of existing methods of estimating *m* while proposing and assessing a cross validated method for estimating *m* across various settings. These methods were then considered in a study incorporating EFA to assess the relationship and categorization of self-reported chronic rhinosinusitis (CRS) symptoms, a common sinus inflammatory disease, within three cross sectional questionnaires as well as within the in the changes in symptoms between questionnaires.

A cross validated approach (trace) was developed by which *m* increases until the discrepancy between the implied correlation of a partition of data and the observed correlation of the other data partition increases. In order to assess the performance of this new method as well as other, common approaches, a simulation study was designed in which valid factor loading matrices were simulated using a new procedure, and random samples were drawn from their respective correlation matrices. The trace method displayed quickly increasing accuracy when more samples were drawn, a phenomenon not observed in other methods. Trace was also applied to the CRS data, suggesting 13 factors to be extracted, more than other methods. This non-agreement possibly highlights the differences in factor extraction interpretations, and the different meanings of "correct" *m*.

An EFA was carried out on self-reported CRS symptoms as well as changes in symptom responses over time in order to identify any relationships between or categorization of CRS

symptoms. A total of 3535 primary care patients were included this study having responded to three questionnaires of 37 repeated questions spanning a 16-month period. After extracting factors from all three questionnaires and two symptom difference scores, five stable factors were identified in each. The factors of congestion and discharge, facial pain and pressure, smell loss, asthma and constitutional as well as ear and eye symptoms were consistent with the hypothesis that CRS symptoms are measuring several distinct biological processes.

Readers: Karen Bandeen-Roche, Brian Schwartz

Acknowledgments:

I would first like to thank my thesis advisors and readers Dr. Karen Bandeen-Roche and Dr. Brian Schwartz of the Johns Hopkins Bloomberg School of Public Health from whom I have learned so much. With their guidance and support I was allowed to work and develop ideas on my own while being steered back on track when needed.

I must also express my gratitude to my parents who have supported and encouraged me through my studies. Without their support, this work would not be possible.

Table of Contents

List of Tables

Chapter 2:

Chapter 3:

List of Figures

Chapter 2:

Chapter 3:

Chapter 1 - Introduction

Exploratory factor analysis (EFA) is a statistical method utilized to investigate and summarize the joint distribution of a collection of variables through the estimation of the relationship between these observed variables and unobserved but theorized factors. It relies on the assumption that covariance among measured variables arises from a smaller set of latent factors which are associated, to varying degrees, to each observable variable (known as the common factor model assumption). These methods are commonly utilized when there is little to no *a priori* knowledge about the latent structure associated with variables, and have been employed across a variety of scientific domains. Commonly, EFAs are employed in an attempt to better understand related phenomena, to allow researchers to create scales, and as an intuitive way to study what a collection of observations is measuring.

Despite being a long-established technique, considerable difficulty still is encountered when employing this approach. The most common issue practitioners face while attempting to utilize an EFA is which factor model to utilize—largely, how many factors to incorporate. Determining the number of factors to include (m) can be a difficult problem as a subtle change in m can vastly impact the results and interpretations of the analysis, and the choice of m itself can be very ambiguous. Furthermore, the problem of selecting m has been approached from a variety of angles, none of which has earned consensus agreement as a gold standard method. Part of this thesis includes work studying existing methods for choosing the number of factors (m), while proposing and examining a new method which incorporates a cross validated approach to selecting m.

In addition, we applied the principles studied to an analysis of chronic rhinosinusitis (CRS) symptom data collected from patients identified in a large, integrated health system. A total of 3535 patients responded to 37 questions pertaining to a spectrum of CRS symptoms three times (baseline, 6-month and 16-month follow-up questionnaires). The CRS study, in fact, motivated the statistical work. EFA was conducted to better understand relationships among and categorization of CRS and CRS related symptoms. EFAs were conducted for each of the three questionnaires as well as for the change in reported symptom frequency (baseline to 6 months, and 6 months to 16 months). We were able to identify five similar factors in each of these analyses, each with a biologically plausible pathological explanation, suggesting that there may be real phenomena driving these observations. At the same time, the selection of five as the factor number was better evidenced in some periods than others, and by some methods than others, illustrating the challenges studied in our first paper.

This thesis comprises two papers, one that utilized EFA in order to explore the covariance structure of self-reported symptoms related to CRS and common, co-morbid conditions, while the other studied the various methods of determining the number of factors in EFA settings and proposed another method to do so. It begins with the methodological work and then proceeds to the detailed CRS analysis. These papers work in tandem by examining the theory and challenges from an analytic and philosophical standpoint of selecting the “optimal” number of factors, while estimating the number of factors and putting EFA to work in a real-world setting involving a common disease. A concluding chapter provides synthesis and identifies areas for future work.

Chapter 2 - A Cross-Validated Approach to Exploratory Factor Analysis Model Selection

Introduction

Frequently it is of public health interest to characterize attributes of data that cannot be measured directly. Instead, a group of observable variables that indirectly characterize the unobserved are collected. In such settings, it is often of importance to explore the underlying, "latent" structure of data in addition to the observed manifest variables. Ideally, in doing so we can study unobserved, latent variables which may be more interpretable or of a greater importance than their measured counterparts. One method of estimating latent structure is through factor analysis (FA). This method is common in diverse fields ranging from psychology and economics to health and spirituality (Fabrigar, Wegener, MacCallum, & Strahan, 1999; Hirose, Kawano, Konishi, & Ichikawa, 2011; Underwood & Teresi, 2002).

There are two general cases of factor analysis, exploratory and confirmatory. Exploratory factor analysis (EFA) aims to identify the underlying structure of variables with little *a priori* knowledge of any such relationship, while confirmatory factor analysis is utilized to test whether a proposed latent structure adequately fits the observed data. While both are useful and powerful techniques, exploratory factor analysis requires the additional step of choosing $\hat{m}$, the estimated number of factors (m) characterizing the observed data distribution, a model selection problem which will be considered below. This model selection is important as both

the quantitative and qualitative results of an EFA may rely heavily on this selection, such that different specifications of m may potentially lead to altered interpretations and inferences. Current methods of determining m vary with respect to computation as well as theory, some utilizing likelihood based methods, including Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), to assess the hypothesis that the data is generated from models with specific m, while others utilize properties of correlation matrices (in terms of eigenvalues), or other criteria to test model fit. Some methods of estimating m are useful, but most have drawbacks as well.

The FA approach is based on the common factor model by which observed variables are conceptualized to arise from 3 components: common factors, unique variability, and measurement errors / noise (Brown, 2014). The core equation of the common factor model is as follows in scalar notation:

$$y_{i,j} = \sum_{g=1}^{m} \lambda_{j,g} \nu_{i,g} \; + \; \epsilon_{i,j}$$

where $y_{i,j}$ is the value of variable j for person i, $\nu_{i,g}$ is the g^{th} factor variable for person i, $\lambda_{j,g}$ is the "loading" of the j^{th} variable onto the g^{th} factor, and $\epsilon_{i,j}$ is the residual term for person i and variable *j* which remains unexplained by the factor model specific to our jth variable, the sum of unique natural variation and measurement error (Brown, 2014). Without repeated observations, the noise term is not distinguishable from the natural variation term, and in the above equation, both comprise $\epsilon_{i,j}$. The factors impact, to varying degrees as reflected by their loadings, many observed variables.

It was our aim to introduce a cross validated approach by which m was chosen to be the value which reduced the difference between distributions implied by an estimated factor model and empirically characterizing independently held out data.

Factor analysis model selection is difficult and commonly relies on qualitative or biased methods. There is a need to explore alternative methods of identifying m which display desirable properties, such as close proximity to an underlying data distribution or reproducibility. We also perceive need to evaluate methods in settings in which the number of observed variables is large relative to the sample size. As elaborated shortly: Current methods can be inaccurate across some or all testing attributes (e.g. correlation strength, factor structure, and sample size). The proposed method utilizes a cross-validated approach to determine when the addition of a factor does not summarize additional common variability in the EFA model. In more precise terms, the proposed method continually increases the number of factors (increasing $\hat{m}$) until the difference between the observed and proposed correlation matrix, as measured by a discrepancy function, increases. Cross-validation helps us avoid the following pitfall: As the number of factors (m) increases towards the number of variables, p, in a single sample, the 'difference' between the observed and implied correlation matrix will decrease - even if only due to noise. By incorporating a cross validated approach, we expect that the addition of a factor that only characterizes noise in one partition should actually worsen fit in another. When this phenomenon is observed, we propose that the previously utilized number of factors is a better fit for the data at hand.

This study was motivated by a project investigating chronic rhinosinusitis (CRS), a sinus inflammatory disease impacting approximately 15% of the United States adult population (Tan, Kern, Schleimer, & Schwartz, 2013). CRS is commonly defined by the presence of 4 cardinal symptoms associated with sinus swelling presiding for an extended period of time, but its diagnosis typically also requires objective evidence of inflammation such as by computerized tomography (CT) scanning. One barrier to the effective diagnosis and treatment of the disease is that the connection between objective inflammation and patient symptoms of sinus opacification is not well understood (Wj et al., 2012). Additionally, obtaining objective evidence can be difficult in resource-limited or high-volume settings, making an improved symptom-based method of diagnosis desirable. To begin addressing these barriers, the motivating study assessed a large sample of patients from a large, integrated health system for presence and severity of a large number of sinus-related symptoms. EFA was proposed as a method to summarize symptom clustering and hence facilitate the subsequent study of symptom relationships with objective evidence of CRS (Cole, Schwartz, & Bandeen-Roche, 2017).

In the remainder of our paper, we examine existing methods for estimating the number of factors in EFA settings as a background for this study, discussing some previously established strengths and weaknesses of each. Then we introduce the proposed method and provide a simulation study comparing efficacy of methods across potential circumstances including sample sizes, correlation strengths and distributions, as well as number of variables. Results are compared across methods, correlation structures, and sample sizes. Finally, we apply the various methods for selecting the number of factors to the CRS study. The CRS findings themselves are presented in the next thesis paper.

Background

Notation and Assumptions

Exploratory factor analysis aims to represent a multivariate data distribution (commonly through the correlation matrix, R) according to the common factor model. This model is characterized by parameters we collectively label as "θ", consisting of a $p \times m$ loading matrix, Λ, $p \times p$ inter-factor correlation matrix Ψ, and $p \times p$ unique variance (diagonal) matrix, Δ_Ψ. Because each θ "implies" exactly one correlation matrix, $P(\theta)$, which is the model's characterization of the matrix of correlations among the observed variables, we can make statements about an EFA's implied correlation matrix which has the form $P(\theta) = \Lambda\Psi\Lambda' + \Delta_\psi$.

The common factor model represents the measured variables as functions of latent (unobserved) factors as well as model parameters, most notably factor loadings. In matrix notation (scalar representation has been previously provided) $\boldsymbol{y_i} = \Lambda\zeta + \delta$ where $\boldsymbol{y_i}$ is the $(p \times 1)$ vector of observed variables for person i, Λ is, again, the $(p \times m)$ loading matrix, ζ is the $(m \times 1)$ latent factor vector, and δ is the $(p \times 1)$ vector of error terms for each individual. Each element δ is assumed to be independent from each other and independent of ζ as well.

Common assumptions of this model are that the error terms are mutually independent and independent of the factor variables, and that the collection of the factor and error term variables is multivariate normally distributed. The variances of variable-specific residual terms, Var(ϵ_j), are referred to as "uniqueness" terms that remain unexplained by the common factors, leaving Var(Y_j)-Var(ϵ_j) as the "common" or shared variance ("commonality") in a given observed variable that is attributable to its factor contributions.

For the remainder of this paper, N will denote the sample size.

EFA and PCA

It may be worth making a quick note of the difference between EFA and principal components analysis (PCA), two related and commonly confused, yet distinct, techniques. The aim of PCA is to reduce the dimensionality of data while retaining as much information (variability) as possible through the conversion of correlated variables into a set of uncorrelated linear combinations of (often standardized) observed variables called principal components. EFA is a model-based analysis that aims to identify the relationship between hypothesized latent, but potentially related factors, and the observed variables. While both of these methods are dimension reduction techniques, each is used to answer very different questions, and are not interchangeable.

Existing EFA Model Selection Strategies

There are many common approaches to EFA model selection, each selecting an $\hat{m}$ based on some criteria that try to identify an 'optimal' number of factors, m. If an appropriate choice for m exists, choosing $\hat{m} > m$ is called overfactoring. Overfactoring may result in a misunderstanding of real latent constructs present in data as true factors may be superfluously split into several factors or factors which have 'random' low loading variables may appear (Norris & Lecavalier, 2010). On the other hand, choosing $\hat{m} < m$ is called underfactoring. Underfactoring is considered to be a more serious concern, as observed factors will load

erroneously onto factors to which they don't belong, providing misleading evidence for factor identity (Norris & Lecavalier, 2010).

Several strategies for estimating m rely heavily on assessing and interpreting the eigenvalues of a covariance or correlation matrix-either of the observed variables, or that is estimated to arise from the common factor portion of the factor model. For the remainder of this paper we will assume that analyses are based on correlation matrices: this has the advantage of standardizing all variances to equal one and all covariation measures to lie on a -1, 1 range. In either case, the eigenvalue of a factor is representative of the amount of variability the factor contributes to the sum of variable variances in most factoring methods (Norris & Lecavalier, 2010). Computationally, a factor's eigenvalue with respect to the observed variable correlation matrix (when factors are orthogonal) is its sum of squared loadings (Norris & Lecavalier, 2010).

There are graphical methods that employ correlation matrix eigenvalues as well, such as Cattell's scree test (Cattell, 1966). This test consists of examining a plot of eigenvalues versus factor index (1, 2, 3, …) to visually select the number of factors to extract. Ideally, there will be an initial steep drop in eigenvalues followed by a clear leveling out in the trend of remaining decrease, creating a classic 'elbow' shape (Cattell, 1966). The number of eigenvalues before the elbow is the proposed number of factors. While intuitive, a clear problem with this technique is that there is not always a clear 'elbow shape', leaving potential for varying interpretation by different investigators.

One of the most popular methods for estimating the number of factors is the Kaiser test (K1), which is based on $e_1, e_2, \ldots e_k$—the eigenvalues of the observed variable correlation matrix. K1 chooses the number of factors to be equal to the number of eigenvalues greater than 1, $\hat{m} = \Sigma_1^k \, I(e_i > 1)$ (Cattell, 1966; Kaiser, 1960). This method provides an intuitive understanding of factor retention methods by which one retains the factors accounting for more than a single standardized variable worth of variability (Velicer, Eaton, & Fava, 2000). It tends to overestimate the number of components in PCA settings, however, potentially due to random noise pushing 'borderline' eigenvalues over the 'threshold' of 1 (Velicer et al., 2000; Zwick & Vejicer, 1984).

A popular extension of the K1 test is parallel analysis, by which observed eigenvalues are compared—typically, plotted—against a measure of central tendency for eigenvalues as simulated from many randomly generated independent (noise) correlation structures (Humphreys & Jr, 1975; Timmerman & Lorenzo-Seva, 2011). The measure of central tendency may be mean, median, possibly some other percentile, or even a single realization of simulated eigenvalues (Humphreys & Jr, 1975). The factors selected is the number of observed eigenvalues that are greater than the simulated eigenvalues (Horn, 1965). Parallel analysis has been considered to be one of the most powerful and accurate methods of determining the number of factors to extract, and has been shown to display better performance than the K1, scree, and other methods while being relatively easy to understand (Velicer et al., 2000; Zwick & Vejicer, 1984). Parallel analysis is sensitive to sample size however, with increased N commonly resulting in more factors being retained, potentially over an optimal amount (Velicer et al., 2000).

Other methods do not explicitly consider eigenvalues and instead take a more traditional model selection approach. Such methods include evaluating likelihoods and functions of likelihoods (Preacher, Zhang, Kim, & Mels, 2013). The likelihood utilized in factor analysis, once logarithmically transformed, is given as

$log(\mathcal{L}) = -\frac{1}{2}N(log(|P(\theta)|) + tr(P(\theta)^{-1}R))$ when each of the N observations are taken to be normally distributed and independent of one another (Akaike, 1987). Likelihood ratio tests can be created which assess the null hypothesis that the observed data are generated from factor model with a specific m (Preacher et al., 2013). Typically, one continues to increase m until the factor model fits the data (lowest m for which the null is not rejected by the likelihood ratio test). Unfortunately, this method comes with several drawbacks. Large sample sizes cause even 'small' discrepancies between model and observed data to cause a rejection while in small sample size situations, large discrepancies may not be identified, leaving performance to be determined largely by the given sample size (Norris & Lecavalier, 2010).

In addition to likelihood ratios, various extensions including the AIC and BIC as well as BIC derivatives such as sample size adjusted BIC (SSBIC) and empirical BIC (EBIC) have been frequently utilized in a factor analysis model selection framework (Hirose et al., 2011; Lopes & West, 2004; Press & Shigemasu, 1999).

For the k orthogonal-factor model, $AIC(m) = -2\times log(\mathcal{L}(m)) + [2p(m+1) - m(m-1)]$ (Akaike, 1987), $BIC(m) = -2\times log(\mathcal{L}(m)) + \log(n)[p(m+1) - 0.5m(m-1)]$ (Lopes & West, 2004), and $SSBIC = -2\times log\big(\mathcal{L}(m)\big) + m \times \log({}^{n+2}/_{24})$ (Sclove, 1987), where m is the number of factors, p is the number of observed variables and n is the number of

observations utilized (Akaike, 1973). Although closely related, these methods can produce very different estimates of $\hat{m}$ (Hirose et al., 2011). This is not surprising, because AIC is directed toward optimizing prediction, whereas BIC was designed to identify "true" model complexity (Akaike, 1973; Schwarz, 1978). In addition, other extensions of AIC and BIC have been produced, and applied in the context of factor analysis including methods such as generalized BIC (Hirose et al., 2011).

Bootstrapping methods have been proposed to provide an alternative method of determining the number of factors while producing a measure of uncertainty (Thompson, 1988). Some such procedures draw bootstrap samples and then estimate m for each sample using a common approach such as parallel analysis: by doing this, one could obtain a bootstrap interval for m, which could inform an appropriate range of values for m. In similar fashion bootstrap intervals for commonality, loadings and inter-factor correlation measures also could be provided. Other, recent methodological work has shown the efficacy of cross validated methodologies as well. A "bi-cross-validation" technique proposed by Owen & Wang (2016) randomly holds out submatrices of the data matrix, against which factor model predictions developed on the remaining components of the data matrix are tested. This method has been shown to outperform a variety of other methods, even parallel analysis, under certain simulated situations (Owen & Wang, 2016). This method evaluates predictions with respect to a submatrix of the observed data matrix, in contrast to the method we shortly propose, which evaluates predictions with respect to a correlation matrix based on a subset of the sampled observations. It also uses a distinct, multi-step procedure to develop predictions and has a

distinct goal of recovering the underlying factor prediction, rather than a "true" number of factors, *m* (Owen & Wang, 2016).

Novel Method for Model Selection

In the ordinary least squares method for estimating the factor analysis model, the discrepancy between and observed and factor-implied correlation matrices can be measured by the following discrepancy function (Lee, Zhang, & Edwards, 2012):

$$f = \frac{1}{2}\text{TRACE}((R - P(\theta))^2),$$

where the trace function is the sum of diagonal matrix entries. We can assess factor model fit by tracking the value of f at varying numbers of factors. If fitting and assessment are conducted within one, same dataset, we expect that f will improve (decrease) as m, the number of factors, increases. When cross-validation is applied, however, with fitting conducted in one subset and testing applied in another, we expect that f will improve as m increases to a point, but then worsen as one exceeds the dimension needed to characterize the true data distribution. The specific procedure we propose is as follows:

Procedure:

1. Randomly split data into a training and testing set ($h = 2$)
2. Calculate correlation matrix for training and testing set

3. Fit factor models (using m from 2 to p) using the training matrix
4. Compute the value of trace function comparing the implied correlation matrix from the training data to observed data test correlation matrix.
5. Find where f increases, stop there. Our choice ($\hat{m}$) is the last 'step' before f increases. If f increases indefinitely, our best estimate of $\hat{m}$ will be p, indicating that a factor model is not a parsimonious fit for the data at hand.

Simulation Study

We assessed the effectiveness of our proposed method relative to common existing model selection methods in a simulation study. Random samples from factor models with known m were generated and the proportion of samples for which $\hat{m} = m$ was estimated for a variety of correlation structures arising from different factor models. Factor models were fit using the ordinary least squares (OLS) method as implemented by the psych R package (Lee et al., 2012; R Core Team, 2016; Revelle, 2017). It was our aim to compare findings over correlation structures varying in several different aspects including: strengths of loadings, number of observed variables, and distribution of loadings among factors. The procedure outlined below was utilized in order to provide loadings/correlation matrices with both structure and elements of randomness in the hopes of approximating real-world situations.

Nine factor structures were utilized, six from a simulated theoretical framework and three incorporating the implied correlation matrices from the CRS EFA which motivated this

study. EFA structures for five-factor models (representing assessments at baseline, 6 month follow up, 16 month follow up), were utilized in order to provide structure to this simulation study which approximated a complex empirical scenario, while being determined in advance and thus feasibly capable of model selection method accuracy. Out of the six theoretical matrices, three "blocked" matrices (named weak, moderate, and strong) contained 25 variables and 5 factors, with each factor having exactly 5 variables loading onto it weakly, moderately, or strongly (mean loadings = 0.38, 0.56, and 0.71 respectively) and remaining loadings only minimally (mean loadings = 0.15, 0.11, and 0.07 respectively). A "moderate, low dimensional" matrix represented a model including 5 factors and 11 observed variables, with 3 variables loading moderately (mean loadings = 0.56) onto the first factor while the remaining factors contained 2 unique variables loading moderately; minimal loadings were 0.11 on average. A "moderate, different dimensional" matrix contained 5 factors and 27 variables, with factors loading on 10, 7, 5, 3, and 2 variables respectively with mean loading of 0.56 while minimal loadings were 0.11 on average. A 10-factor, 100 variable matrix was also utilized with 10 variables loading heavily onto each factor (mean loadings = 0.037) and 90 variables loading minimally (mean loadings = 0.007).

Theoretical matrices described above were generated as follows. By treating each of p rows of the loading matrix Λ as an m dimensional Dirichlet vector with parameters $\alpha_{i,1}, \ldots \alpha_{i,m}$ we can generate a valid loading matrix for which the sum of squares for each row is less than or equal to 1 (avoiding a Heywood case), and each factor matrix entry is between -1 and 1. We can then compute the commonalities as $\Delta_\Psi = I - diag(\Lambda\Lambda')$. The matrix: $\Lambda\Lambda' + \Delta_\Psi$ represents our simulated correlation matrix generated from a known factor structure: Figures 1 and 2 illustrate

the result for the strong “blocked” correlation design, and others are illustrated in the appendix. This factor structure is completely determined by the choice of p, m, and all Dirichlet parameter values. Although this approach is simple, it is capable of generating a wide range of structures (see appendix).

For each correlation matrix, simulation runs were conducted for sample sizes of *N* = 100, 300, 500, 700, 1000. In each run, samples of the respective size (*N*) were drawn from a multivariate normal distribution with mean zero and the respective correlation matrix as its covariance matrix. Using these simulated samples, K1, parallel analysis, BIC, sample size adjusted BIC, and the proposed method were applied to determine the number of factors. For each run, 100 repetitions were conducted, and the number of successful estimates (alternatively, the percent of correct estimations) were recorded.

Results

The proposed trace function method performed increasingly well with increasing sample sizes. In some cases (e.g. 'moderate' blocked structure) accuracy increased from approximately 5% at $N = 100$ to 100% at $N = 1000$, while the difference in accuracy of the other, standard methods varied less across N (**Table 1**). Weaker structures were more difficult for the proposed method to identify. For the 'moderate, low dim' matrix, accuracy varied from 5 to 35% as one increased the sample size from 100 to 1000 (**Table 1**).

While the trace function improved its estimation from 100 samples to 1000 for each proposed correlation matrix, the K1 method as well as the SSBIC method did not once improve

from the same change in *N* (**Table 1**). Meanwhile the BIC method improved only once in the 9 proposed scenarios while the parallel analysis improved in two of the 9 (**Table 1**). Overall, all tested methods performed very well with stronger correlation structures, although BIC performed comparatively worse in the CRS and 10-factor settings (**Table 1**).

There were cases where standard methods performed outstandingly well. In the strong correlation simulations, the SSBIC, K1, and parallel analysis approaches performed at 100% accuracy across N. Simulations involving the three CRS correlation structures and the ten factor structure saw perfect accuracy across *N* with SSBIC and K1 methods while the trace method achieved at least 97% accuracy at *N* of 1000 (**Table 1**). In the moderate low dimensional and moderate different dimensional matrices at low *N*, all methods performed poorly, with the trace function (*N* = 100, accuracy = 5%), and K1 methods achieving the best results (*N* = 100, accuracy = 16%) in each respective scenario (**Table 1**).

Application to the CRS Study

Our analysis addresses 3535 Geisinger Health System patients who were followed for a duration of 16 months, each of whom was selected using a stratified sampling method designed to oversample racial minorities and those with a high propensity for CRS via *International Classification of Diseases* (ICD-9) and *Current Procedural Terminology* defined attributes in electronic health record data (Hirsch et al., 2017; Tustin et al., 2017). Each of these patients responded to three questionnaires containing 37 common questions at baseline, 6 months, and 16 months. Of these 37 questions, 21 inquired about the presence and severity of CRS nasal and sinus while the remaining questions assessed presence of asthma, allergy, ear and

constitutional symptoms. All of the questions inquired about the frequency of experiencing a symptom, or the frequency of being bothered by a symptom in a given timespan and each question was answered on the same Likert scale (1 = never, 2 = once in a while, 3 = some of the time, 4 = most of the time, or 5 = all the time). These questions were specifically designed to predict sinus opacification location and severity using only self-reported symptoms. Polychoric correlations were calculated from each of these surveys, and each of the methods studied in our simulation study was applied to determine the number of factors to extract.

K1 and parallel analysis indicated 6, 5, 7, and 5, 5, 6 factors for baseline, 6 month and 16 month questionnaires respectively while BIC (15, 16, 14), SSBIC (17, 17, 20), and trace (13, 13, 16) suggested substantially higher numbers of factors for the same three questionnaires. Scree plots were also examined, which appeared to suggest 5 factors in each survey. The trace function (Figure 3) may indicate some ambiguity in factor number choice, achieving a minimum at 13 factors, but only modest slope below and above this number – much less than in other simulated scenarios (see appendix). We examined a 13-factor solution for the baseline CRS EFA, extracted using the same OLS method and oblimin rotation as in the 5 factor EFAs, in order to examine the qualitative differences driven by the differences in the number of extracted factors: Two factors' interpretations were invariant--the facial pain and pressure symptom and smell loss symptom factors; other factors however, were reduced to identifying symptoms related to single phenomena or organs (nasal congestion, ear, eye, fatigue, etc.) and addressing both presence and severity (when present in the surveys).

The substantive CRS factor analysis was not approached from a dogmatic factor model vantage point, but rather aimed to identify biologically feasible symptom clusters within the

questionnaire EFAs. As such, scree plots and parallel analysis were prioritized, and in addition was viewed as being parsimonious compared to other methods. Although parallel analysis did not suggest the same number of factors for each of the questionnaires, it was decided to do so for interpretability and comparison reasons. Further elaboration is provided in the CRS EFA chapter.

Discussion

The results of our simulation study provide some insight into the efficacy of the proposed method as well as the effectiveness of other, commonly utilized methods incorporated into this study (K1, parallel analysis, SSBIC, BIC). It was shown that, with increasing sample sizes, the trace function method performed progressively better, eclipsing the performance of other methods in many tested circumstances (**Table 1**). Interestingly, the other methods did not appear to vary in performance with increased N across the majority of tested circumstances (**Table 1**).

In these simulated matrices, the trace function method outperformed the three standard methods when the strength of the underlying factors was small (**Table 1**). The N required to attain superior performance varied, but seemed to be less for weaker correlation structures (**Table 1**). For stronger correlation structures, each of the standard methods tested produced consistently very accurate results (**Table 1**). In all three CRS implied correlation matrices, the sample size adjusted BIC as well as the eigenvalue greater than 1 method achieved perfect accuracy across all values of N, while the standard BIC measure performed

very poorly, with accuracies constantly below 10%. Although the trace function did not attain 100% accuracy across all *N*, it improved from 61, 46 and 34% accuracy at $N = 100$ to 100, 97, and 99% accuracy at $N = 1000$ for baseline, 6 month, and 16 month CRS matrices respectively (**Table 1**).

There have been many methods developed to select the number of factors in an EFA setting, all with slightly different interpretations, strengths and weaknesses. Frequently, practitioners conducting EFA treat the factor model literally and strive to uncover the 'true' m and evaluate their respective loadings accordingly (Preacher et al., 2013). We believe this goal often is unclear, and potentially unhelpful, for two reasons. First, in practice, researchers apply several different criteria to the target they seek, including a literal number of factors underlying the observed data, the most interpretable number of factors, and various other criteria. Second, similarly to linear regressions, models are only approximate - with that in mind we will understand that we will never observe 'truth', only estimations and approximations (Preacher et al., 2013). As such, a factor, once identified, is not necessarily *real* but at best a useful approximation to real (biological or otherwise) phenomenon.

Searching for the literal number of factors is arguably infeasible, outside of extremely controlled settings, because observed variables likely are affected by a huge number of contributing 'factors' (latent or otherwise). Consider the number of factors underlying an individual's take home income. There are likely 'strong' factors including: age, education, work ethic and industry. But there also is a profusion of 'weak' factors such as, appearance, sense of style, and voice quality which may be relatively unmeasurable but could drive stark differences in income amount. Notwithstanding that factor models typically must greatly simplify data

interrelationships, the estimation of factor presence and identity can be an extremely powerful tool, providing insight into disease pathologies and symptom classifications. In the motivating study, EFA was used to better understand the relationships between CRS symptoms and latent factors. It was hypothesized that these factors reflected pathobiological phenomena, shedding light onto what symptom question responses are truly measuring in this population.

As seen in this study's application to the CRS data, differing methods could potentially estimate a wide range of factors. We hypothesize that this discrepancy reflects the differing objective functions employed by the various methods. The K1 and parallel methods are designed to estimate the dimensionality needed to represent shared covariation among one's items, without explicit reference to a factor model. BIC and the trace function both explicitly incorporate the factor model specification in determining fit to the empirical covariance matrix—hence address FA assumptions in addition to dimensionality. It may not then be surprising that these methods require a higher dimensionality to reproduce the empirical data structure. We consider the sensitivity of the dimensionality choice to the factor model assumptions to be instructive in the present case. In the CRS application, simple dimensionality was of interest, rather than consistency with a factor model per se, making prioritization of those methods tuned to this as well as a biologically meaningful interpretation appropriate. When extracting the 13 factors consistent with the trace method from the baseline CRS study, moreover, the majority of factors extracted appeared as 'single symptom' factors, addressing single phenomena or organs. The understanding of these factors is not particularly interesting from a biological or medical standpoint, but may rather suggest 'latent' constructs associated with patient *responses* (similar symptom questions are answered similarly), while other

methods such as parallel analysis provide a more desirable understanding of latent constructs associated with actual patient *symptoms*.

Since the 'true' number of factors *may* have questionable meaning in some circumstances, some have argued that searching for the 'optimal' number of factors may be the *best* course of action, based on a criterion centered around achieving a specific goal such as *verisimilitude* or appearance of reasonable truth (Preacher et al., 2013). Other criteria for "optimal" numbers of factors may include generalizability, the ability to attain similar results on an independent data collected from the same population (Myung, 2000), or accurate and precise data approximation (Owen & Wang, 2016).

The method presented here seeks to approximate an underlying number of factors without forgoing generalizability. A frequent problem in the EFA framework is that EFAs from one sample may not necessarily match another EFA carried out on an independent sample from the same population. By incorporating and embracing the idea of cross validation and generalizability from the beginning, there is a possibility that cross validated approaches may remove variability in the choice of $\hat{m}$ resulting in more consistent results across studies (Friedman, Hastie, & Tibshirani, 2001). These conjectures warrant further research as there are many cases and types of data in which EFA frameworks would be utilized.

An interesting observation from our simulations is that standard methods SSBIC, BIC, and K1, seem to produce the same error rate regardless of N (**Table 1**). This empirical result suggests that these methods may not be consistent estimators of m in these tested settings. The trace function on the other hand nearly monotonically increased in accuracy with

increasing N (**Table 1**). In addition, by utilizing a h-fold cross validated procedure, accuracy may increase more steeply as a function of N as $h > 2$ has been shown to improve estimation accuracy and convergence (Friedman et al., 2001). Several studies have sought to show consistency in the EFA framework, as traditional methods such as BIC may not be consistent for m in all settings (Bai & Ng, 2002). Utilizing a greater range of N here may enable us to show some empirical consistency with some of the methods used, although only the trace function method appeared to approach this across a variety of sample and correlation structure settings.

Future work

Our simulation study provides some insight into the performance of our cross validated discrepancy approach to EFA model selection. However, only a small sampling of possible correlation structures were utilized, all with similar true m ($m = 5$ or $m = 10$). Although we hypothesize that the sign of any loading will not change the accuracy of any of the methods utilized, it is important to note all of the simulated matrices contained only positive factor loadings.

Expanding the proposed method to implement h-fold cross validation is a logical next step. Currently, the data is split into two and *m* is chosen to be the number of factors which minimizes the discrepancy between fit model and held out data. Typically, allowing for an increased number of data partitions leads to more accurate estimates of the error rate, and in this case, may produce more stable estimates of m potentially at lower values of N as well (Friedman et al., 2001).

Additionally, it is important to assess the estimation bias in all methods discussed. Currently, accuracy was assessed only as whether or not the correct number of factors was identified. Considering the importance of understanding a method's tendency to over or under factor, we plan also to assess the magnitude and direction of each method's misses. If over factoring is, generally, better than under factoring for instance, we may wish to penalize overestimation of m less than any underestimation. In addition, missing the number of factors by 1 is likely less deleterious than by 4, for example, so the magnitude by which any given method mis-estimated should be considered in future work.

Tables and Figures

Table 1. Number of correct factor number assessments by sample size adjusted BIC (SSBIC), standard BIC (BIC), Kaiser eigenvalues greater than 1 rule (K1), parallel analysis (PA), and the proposed method (trace) out of 100 simulation replicates.

Simulated samples	**Correlation Structure**	**SSBIC**	**BIC**	**K1**	**PA**	**Trace**
100	Strong	100	98	100	100	79
300	Strong	100	99	100	100	97
500	Strong	100	99	100	100	99
700	Strong	100	98	100	100	97
1000	Strong	100	98	100	100	99
100	Moderate	25	0	92	99	5
300	Moderate	20	0	83	100	69
500	Moderate	23	0	81	100	98
700	Moderate	16	0	90	99	98
1000	Moderate	24	0	86	100	100
100	Weak	2	0	38	75	0
300	Weak	0	0	34	89	30
500	Weak	0	0	44	79	72
700	Weak	1	0	40	82	89
1000	Weak	0	0	33	82	97
100	Moderate, low dim	0	0	0	0	5
300	Moderate, low dim	0	0	0	0	8
500	Moderate, low dim	0	0	0	0	19
700	Moderate, low dim	0	0	0	0	13
1000	Moderate, low dim	0	0	0	0	35
100	Moderate, dif dim	0	0	16	0	2
300	Moderate, dif dim	0	0	14	0	11
500	Moderate, dif dim	0	0	11	0	41
700	Moderate, dif dim	0	0	14	0	54
1000	Moderate, dif dim	0	0	14	0	75
100	CRS ibl	100	7	100	100	61
300	CRS ibl	100	5	100	100	100
500	CRS ibl	100	7	100	99	100

700	CRS ibl	100	3	100	100	100
1000	CRS ibl	100	9	100	99	100
100	CRS i6m	100	0	100	99	46
300	CRS i6m	100	0	100	100	94
500	CRS i6m	100	0	100	98	98
700	CRS i6m	100	0	100	97	95
1000	CRS i6m	100	0	100	99	97
100	CRS i16m	100	0	100	95	34
300	CRS i16m	100	0	100	95	96
500	CRS i16m	100	0	100	91	98
700	CRS i16m	100	0	100	87	96
1000	CRS i16m	100	0	100	91	99
100	10 factor	100	0	100	95	34
300	10 factor	100	0	100	95	96
500	10 factor	100	0	100	91	98
700	10 factor	100	0	100	87	96
1000	10 factor	100	0	100	91	99

Table 2 Estimated number of factors for each of the 3 questionnaires (baseline, 6 month, and 16 month follow ups) from commonly utilized methods including Kaiser eigenvalues greater than 1 rule (K1), parallel analysis (PA), standard BIC (BIC), empirical BIC (EBIC), sample size adjusted BIC (SSBIC), and the proposed method (TRACE).

Method	**Baseline**	**6-month follow up**	**16-month follow up**
K1	6	5	7
PA	5	5	6
BIC	15	16	14
EBIC	8	8	8
SSBIC	17	17	20
TRACE	13	13	16

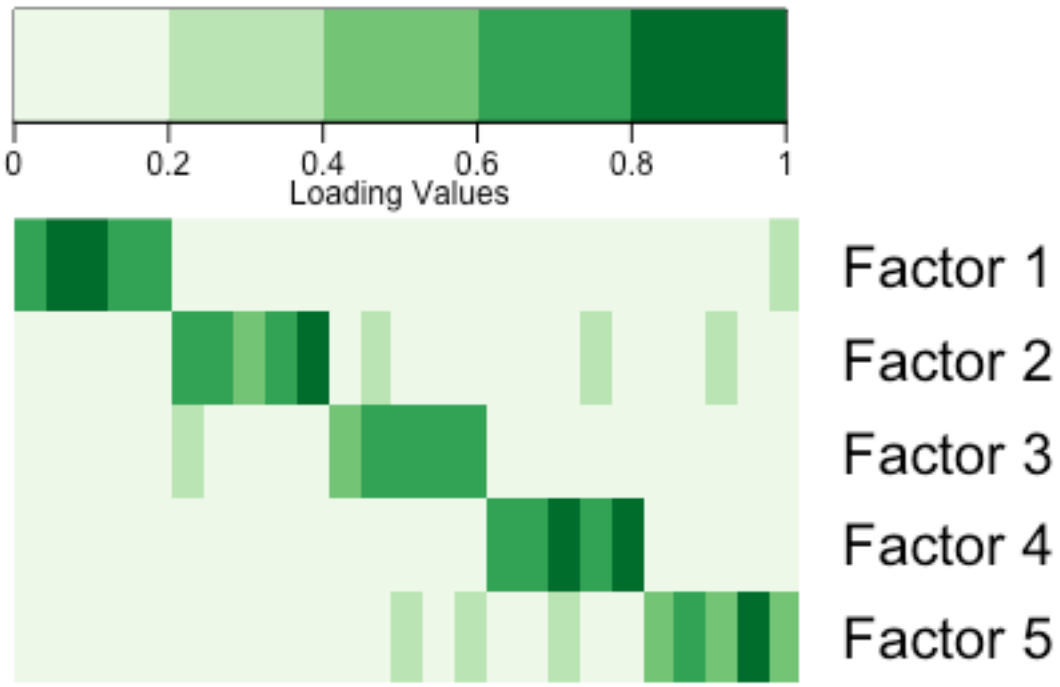

Figure 1. Strong “blocked” factor loadings utilized to create the correlation matrix in the simulation study.

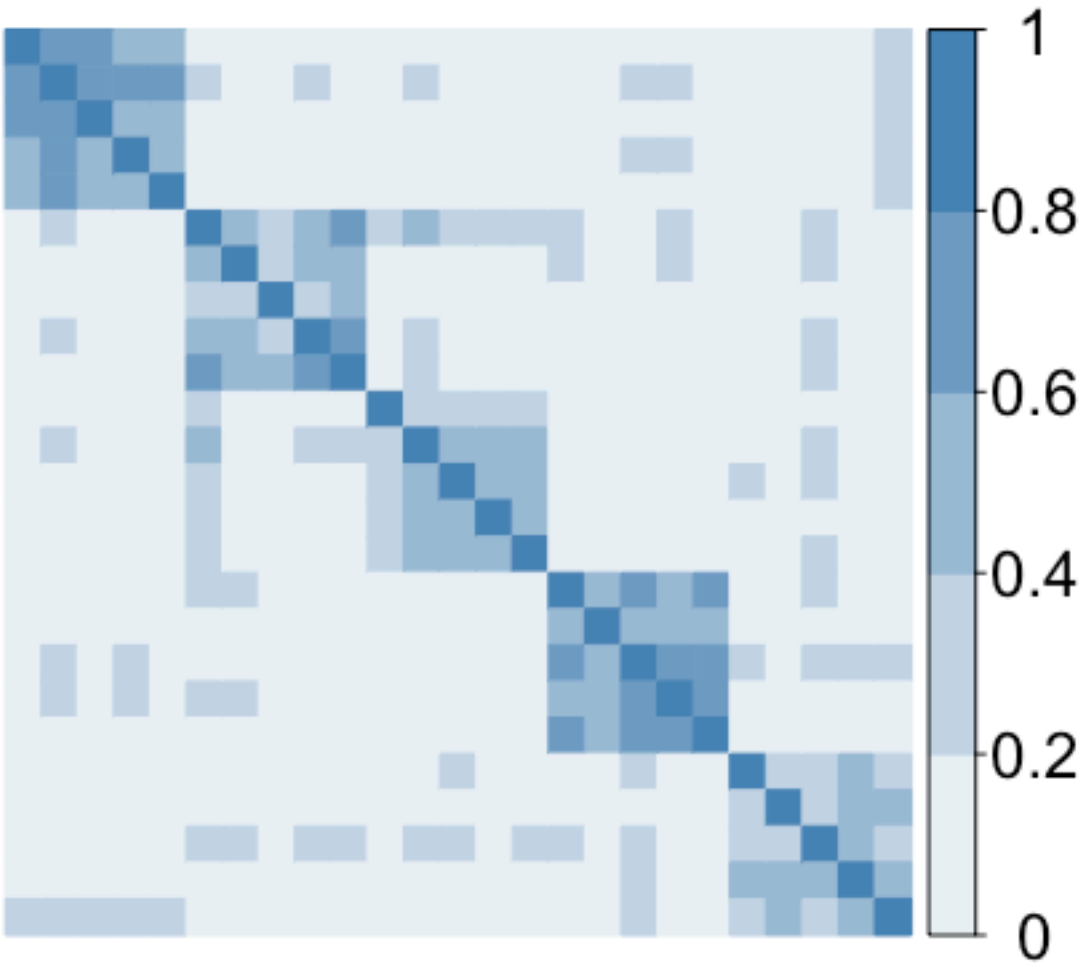

Figure 2. Correlation matrix generated from the strong “blocked” factor loadings matrix.

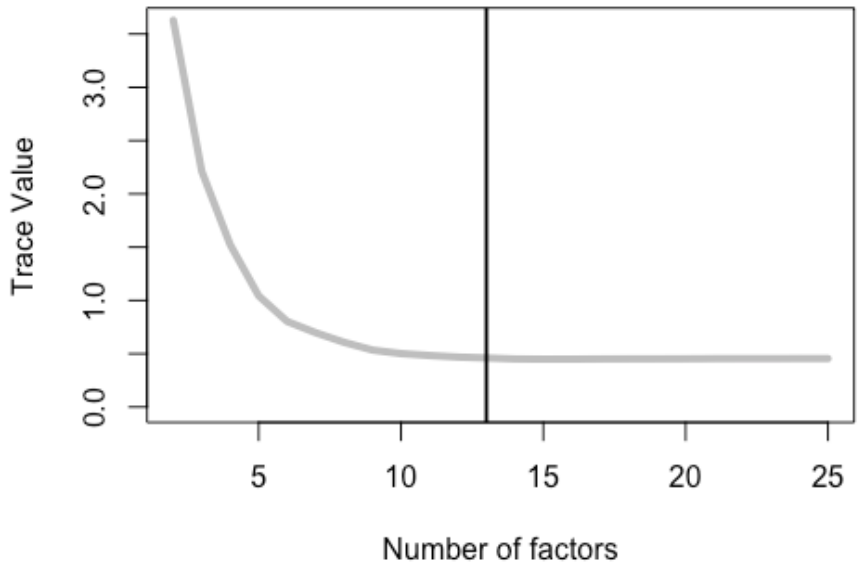

Figure 3 Trace function's discrepancy values on the baseline CRS data. Vertical line denotes the minimum achieved at 13 factors.

Appendix

Additional Figures:

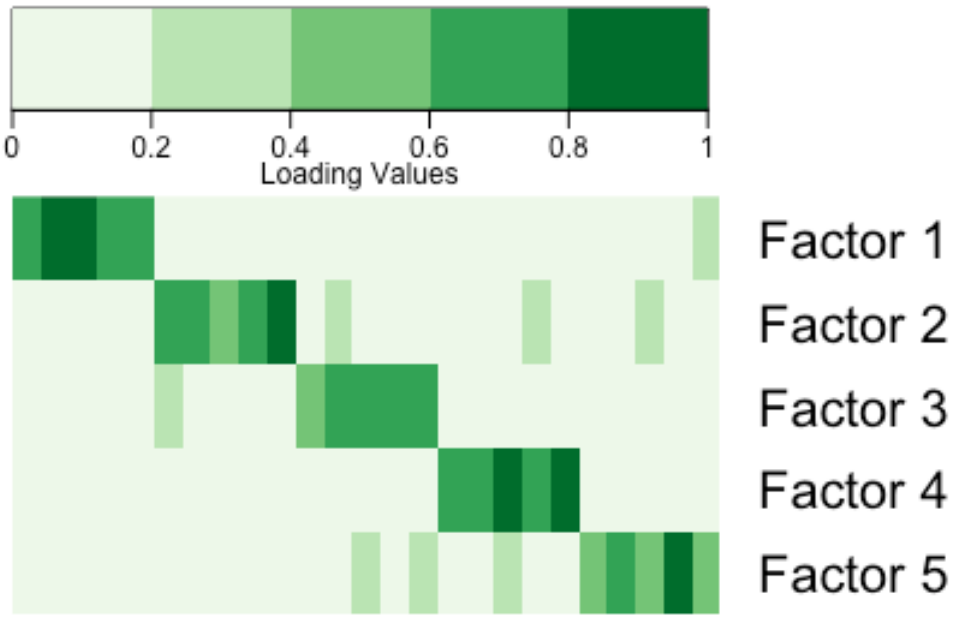

Figure A1. The utilized strong loading matrix, created using the Dirichlet simulation process, consisting of 5 factors and 25 variables.

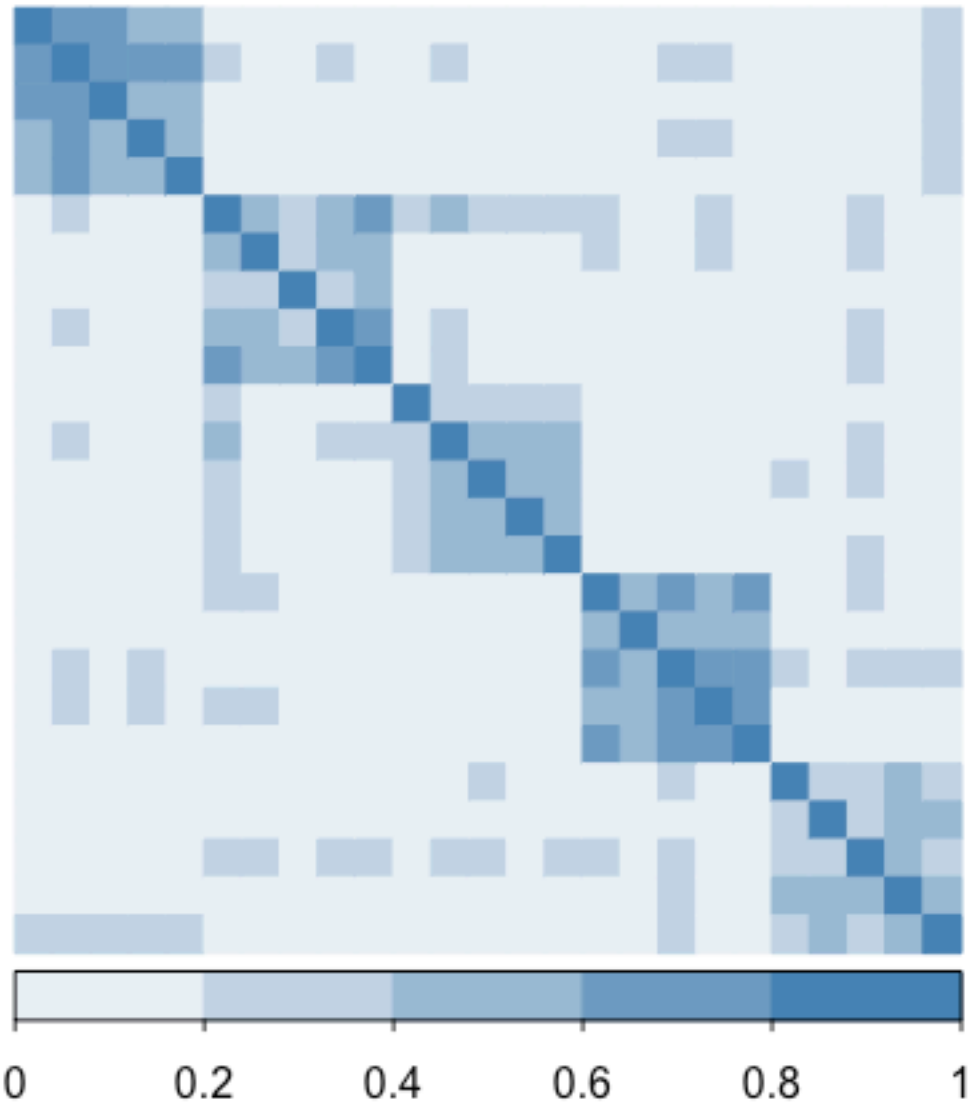

Figure A2. The utilized strong correlation matrix, created from the corresponding strong loading matrix

Figure A3. The utilized moderate loading matrix, created using the Dirichlet simulation process, consisting of 5 factors and 25 variables.

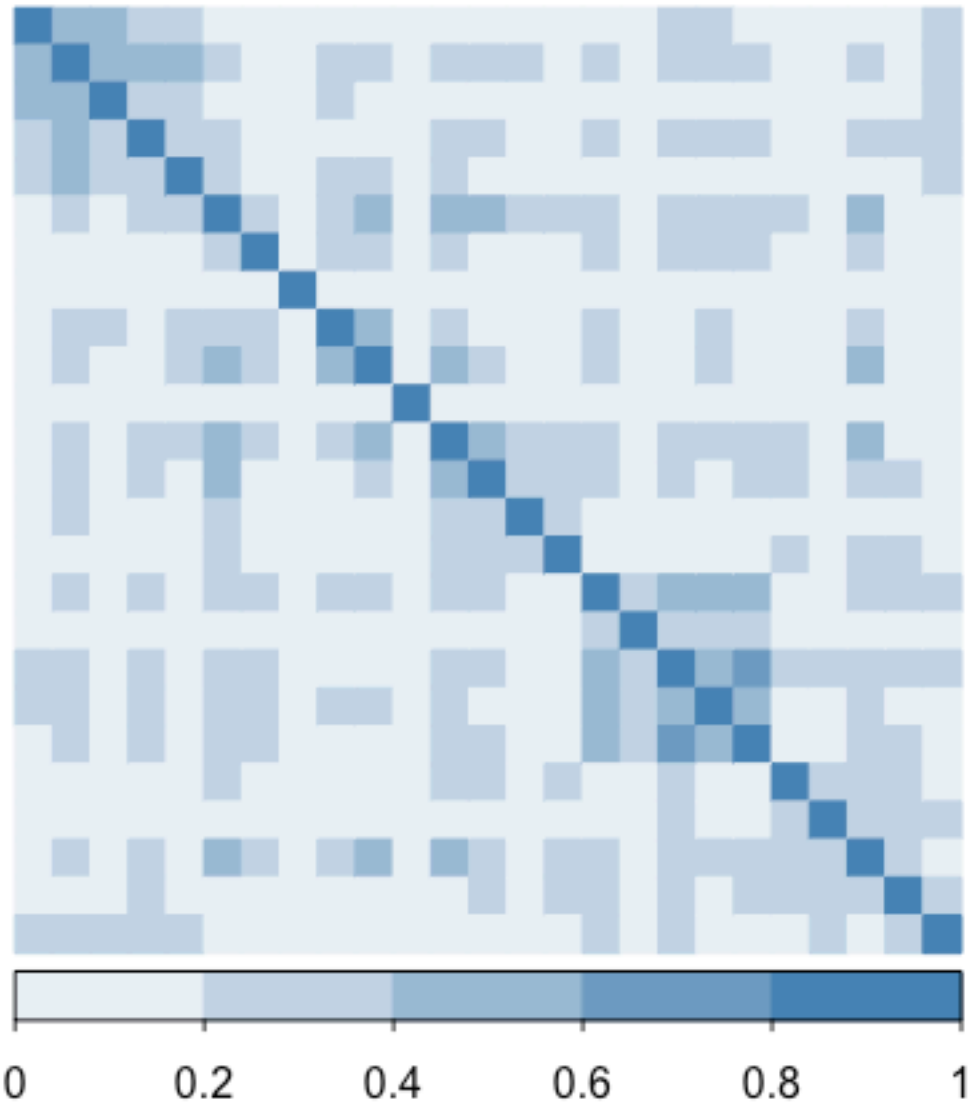

Figure A4. The utilized moderate correlation matrix, created from the corresponding moderate loading matrix

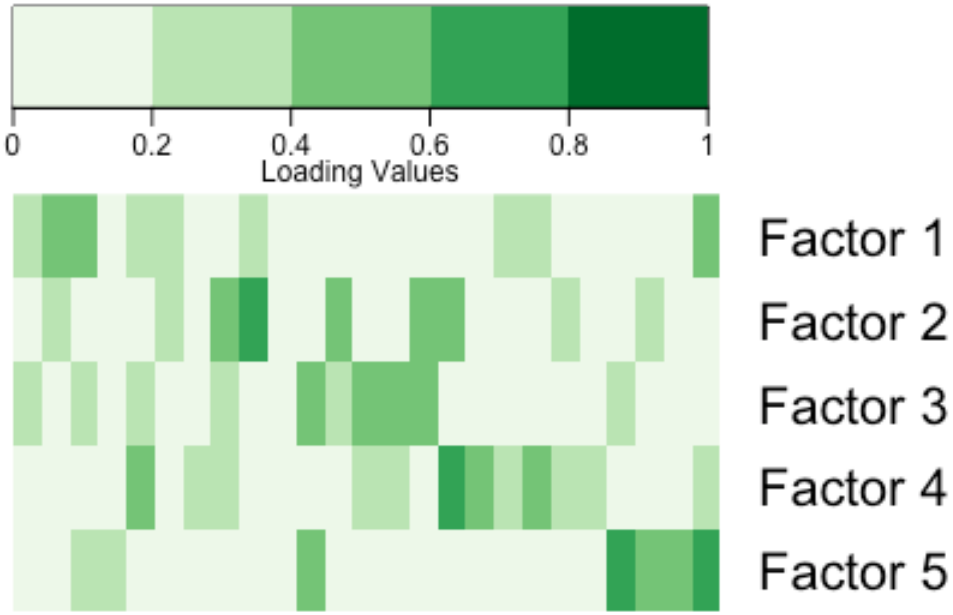

Figure A5. The utilized weak loading matrix, created using the Dirichlet simulation process, corresponding from 5 factors and 25 variables.

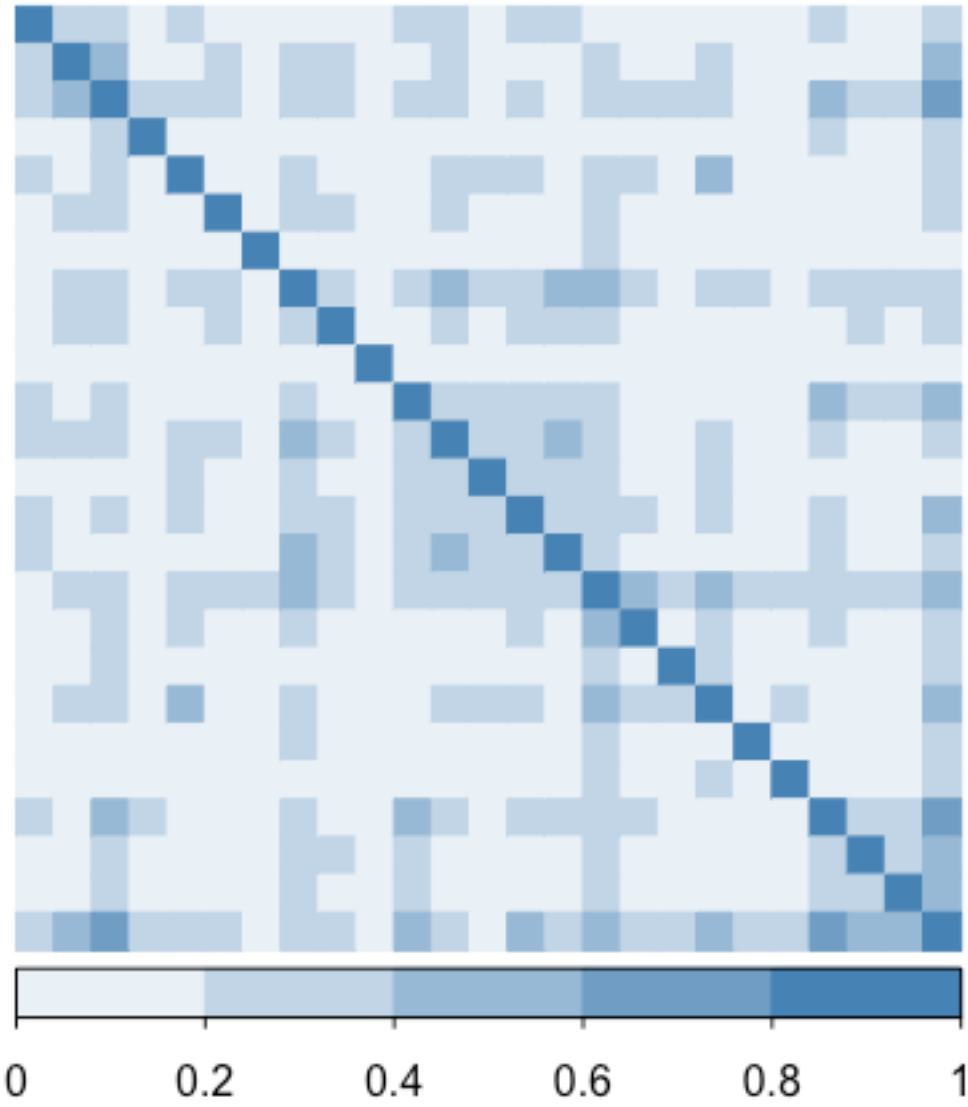

Figure A6. The utilized weak correlation matrix, created from the corresponding weak loading matrix

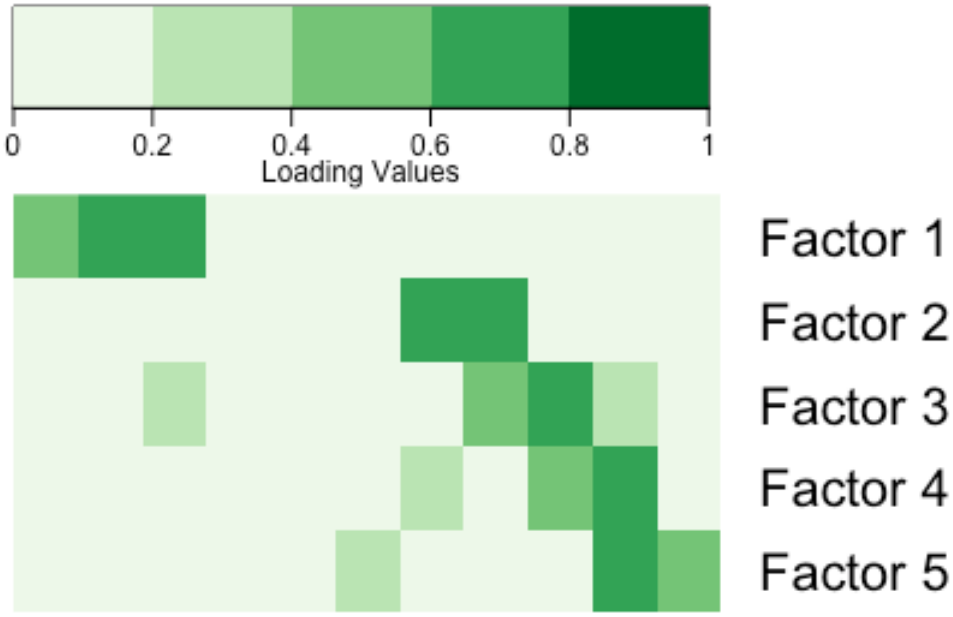

Figure A7. The utilized moderate/low dimensional loading matrix, created using the Dirichlet simulation process, consisting of 5 factors and 11 variables.

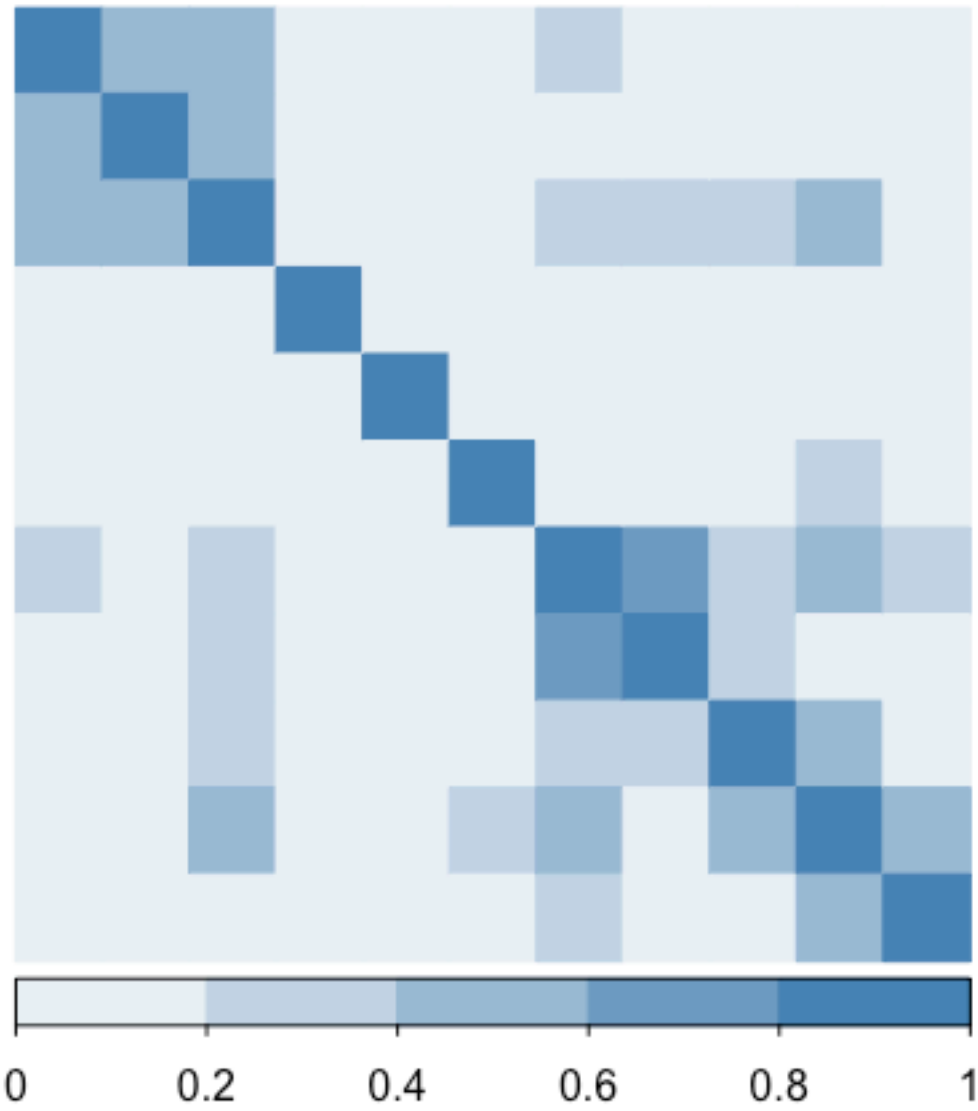

Figure A8. The utilized moderate/low dimensional correlation matrix, created from the corresponding moderate/low dimensional loading matrix

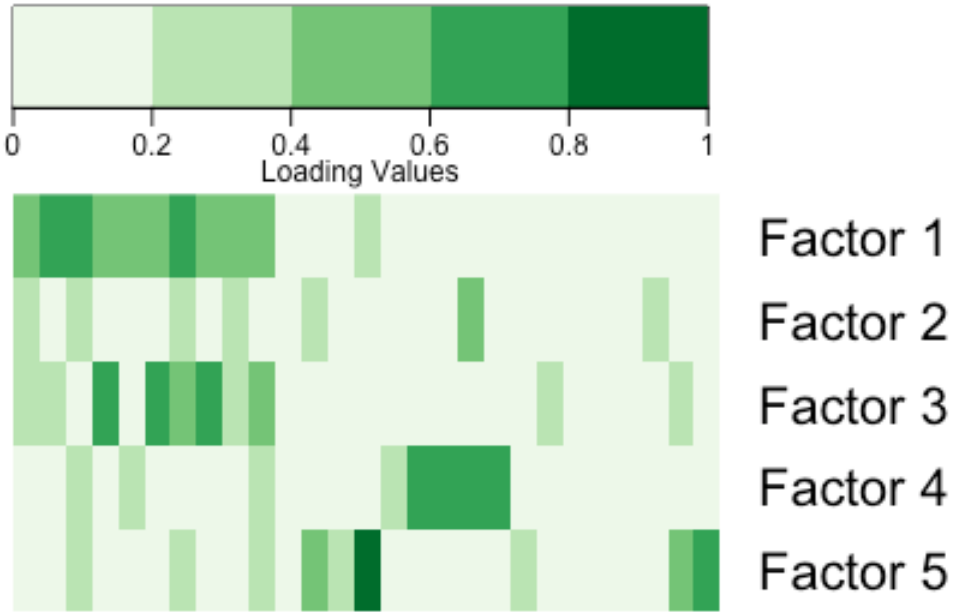

Figure A9. The utilized moderate/different dimension loading matrix, created using the Dirichlet simulation process, consisting of 5 factors and 27 variables.

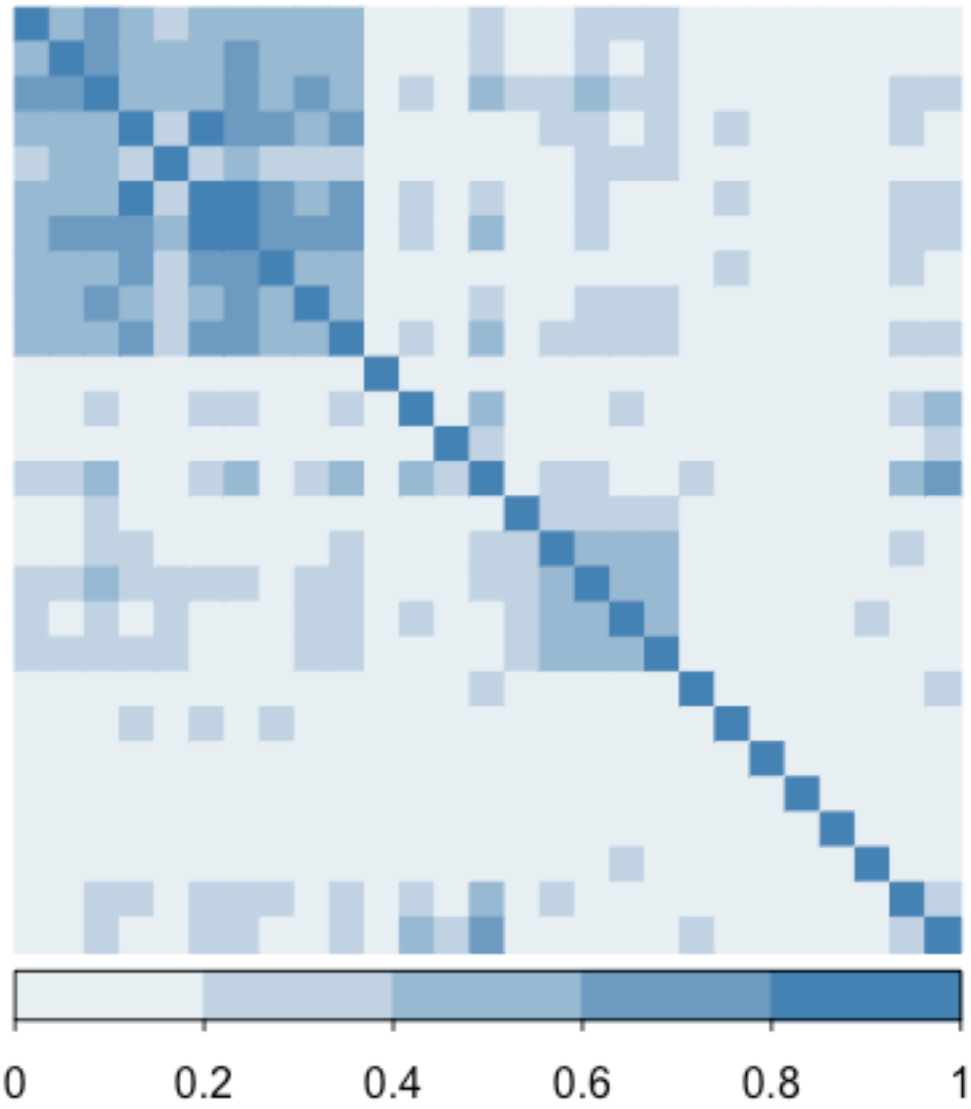

Figure A10. The utilized moderate/different dimensional correlation matrix, created from the corresponding moderate/different dimensional loading matrix

Figure A11. The utilized ten factor loading matrix, created using the Dirichlet simulation process, consisting of 10 factors and 100 variables.

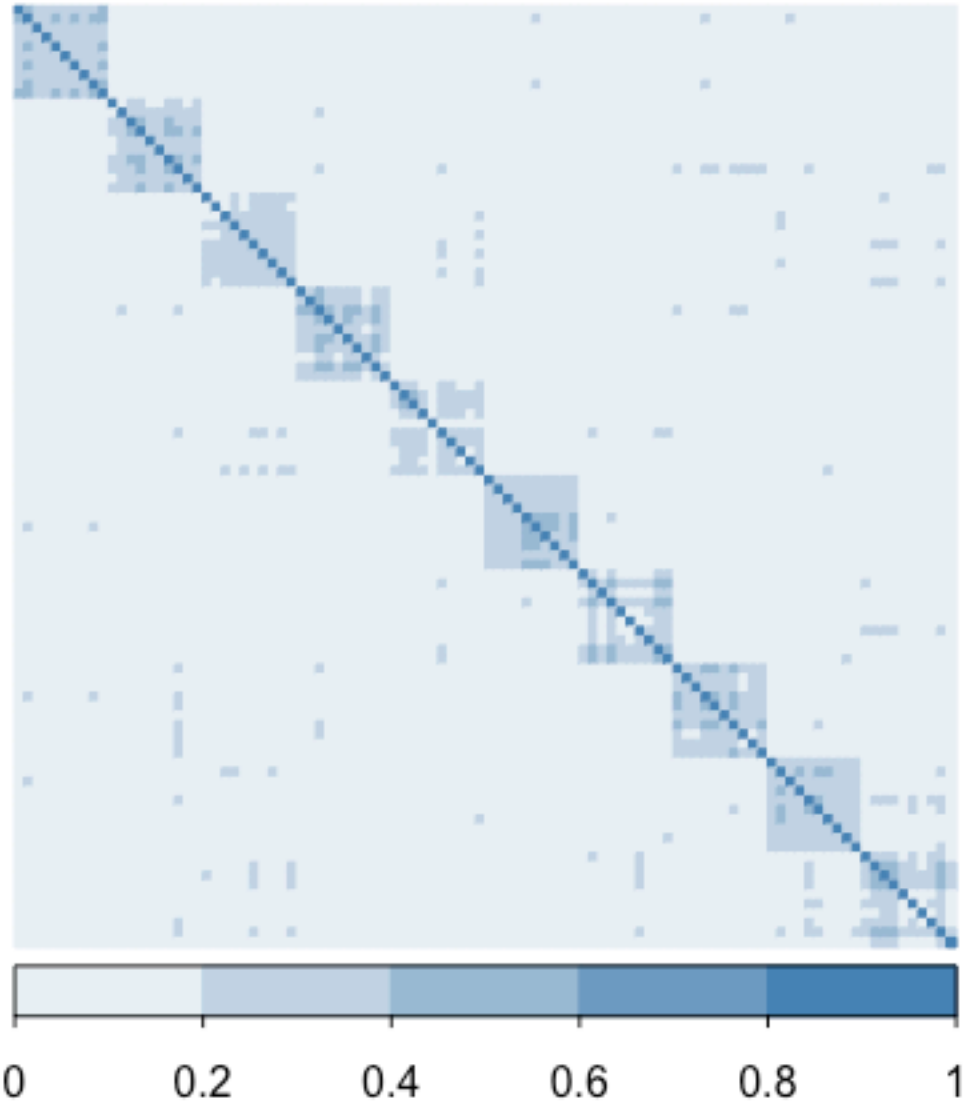

Figure A12. The utilized ten factor correlation matrix, created from the corresponding ten factor loading matrix

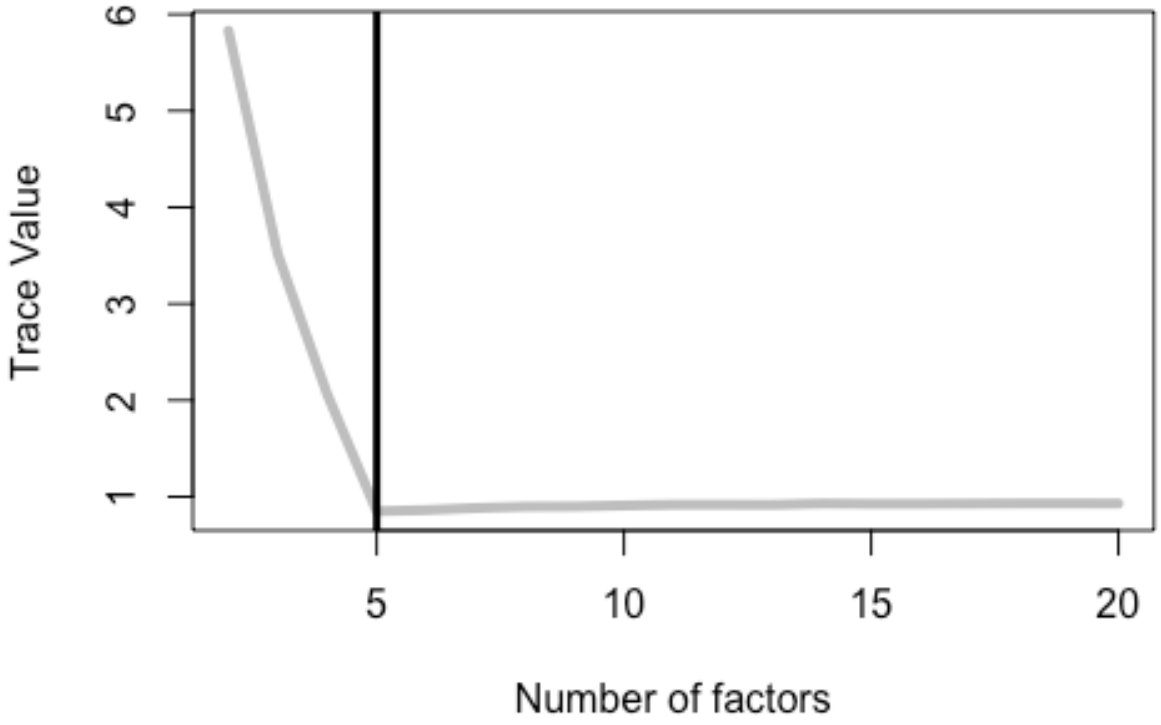

Figure A13. Trace function's discrepancy values on the strong "blocked" factor loading matrix. Vertical line denotes the minimum achieved at 5 factors.

Simulating Factor Model Correlation Matrices

Motivation

In simulations of factor analyses, it is important to be able to randomly generate valid correlation matrices which stem from some known factor model in order to assess model selection methods and other attributes of factor analysis procedures.

Computation

A core idea of factor analysis is that we can explain variability in our observed data by means of a smaller number of underlying, latent factors, which are associated with observed variables. Mathematically speaking, our correlation matrix, ρ, can be broken down as such:

$$\rho = \Lambda\Psi\Lambda' + \Delta_\psi$$

where Λ is the $p \times m$ matrix of factor loadings, Ψ is the $(m \times m)$ factor correlation matrix, and Δ_ψ is the matrix of unique variances ($p \times p$ diagonal matrix).

This can be further written as $\rho = \Lambda\Psi\Lambda' + (I - \text{diag}(\Lambda\Psi\Lambda'))$

If we are interested in generating a random, structured loadings matrix, the following procedure is proposed. We can treat each row i of Λ as a Dirichlet$(\alpha_{i,1}, \ldots, \alpha_{i,m}) \times$ Beta(x_i, y_i) where each $\alpha_{i,j}$ is some proposed weight as to how strong we would like (on average) variable i to load on each factor.

These constraints ensure that Λ will be a valid loadings matrix as:

- All loadings are between -1 and 1
- The sum of squared loadings (for a variable) is less than 1 (avoiding a Haywood case)

Now, we will let ψ be the m dimensional identity matrix (all factors are orthogonal).

$$\rho = \Lambda\Psi\Lambda' + (I - \text{diag}(\Lambda\Psi\Lambda'))$$

$$= \Lambda\Lambda' + (I - \text{diag}(\Lambda\Lambda'))$$

We know $\Lambda\Lambda'$ is positive semi-definite as each element of Λ is a real number.

We also know, a positive semi-definite matrix plus a matrix of the same dimension with all non-negative entries is also positive semi-definite. It follows that ρ is positive semi-definite.

In addition, because each entry of Λ is in [0,1], we know each element of $\Lambda\Lambda'$ will be between [0,1] while the $+(I - \text{diag}(\Lambda\Lambda'))$ term ensures that the diagonal of ρ are all 1. Thus, ρ is a valid correlation matrix, uniquely determined by Λ.

Use

This idea of being able to construct random correlation matrices is important to certain simulation studies where one must generate random yet valid correlation matrices in which the true number of factors is known and fixed. Because any correlation matrix stemming from this method would inherently be able to perfectly decompose into the true number of factors, sampling noise should be added. This can be accomplished by sampling from a multivariate distribution (in this case multivariate normal) with the specified correlation matrix, them computing a 'simulated empirical' correlation matrix from the simulated multivariate data.

Further work

The above method can produce many types of random correlation matrices; however, all loadings must be positive, which is not required of factor analysis loadings. We may be able to incorporate negative loadings by utilizing a Bernoulli process by which each cell of Λ has some probability of being multiplied by -1 or 1. This would allow for negative loadings (and correlations) while ensuring ρ remains a valid correlation matrix.

Chapter 3 - Exploratory Factor Analysis of CRS Symptoms

Introduction

Chronic rhinosinusitis (CRS) is an inflammatory condition characterized by nasal and sinus symptoms, affecting 15% of the United States population (Wj et al., 2012). There are considered to be four cardinal symptoms of the disease which include nasal drainage (anterior or posterior), nasal blockage (congestion), smell loss, and facial pain or pressure lasting for 12 or more weeks (Browne, Hopkins, Slack, & Cano, 2007; Tan, Kern, Schleimer, & Schwartz, 2013). The European Position Paper on Rhinosinusitis and Nasal Polyps (EPOS) diagnosis methodology for CRS is based upon the presence of nasal obstruction or discharge and at least one other symptom as well as objective evidence of inflammation on sinus computerized tomography (CT) scan or sinus endoscopy, which may include sinus or osteomeatal complex mucosal changes, presence of nasal polyps, or mucopurulent discharge from the middle meatus (Wj et al., 2012). Because of the difficulty of obtaining sinus CT or endoscopy in large-scale population studies, EPOS also has an epidemiologic definition of CRS based on symptoms and duration only. However, EPOS does not specify how to measure symptoms in terms of severity (e.g., some blockage or complete blockage; partial smell loss or complete smell loss; the quantity of discharge; the severity of pain) or frequency (e.g., some of the time, most of the time, or all of the time).

Nasal and sinus symptoms lasting three months are quite common, and many studies have reported that there is not a strong correlation between such symptoms and objective

opacification on sinus CT scans (Browne et al., 2007; Ferguson, Narita, Yu, Wagener, & Gwaltney, 2012; Wj et al., 2012). Up to 40% of those with symptoms meeting EPOS criteria for CRS do not have significant sinus opacification on CT (Ferguson et al., 2012). The lack of correlation of symptoms meeting EPOS criteria for CRS and findings on sinus CT could be due to imprecision in the ways that nasal and sinus symptoms have been measured in terms of severity, frequency, and duration (Hamilos, 2011; Wj et al., 2012). In addition, there are few studies that examine how nasal, sinus, and other relevant symptoms relate to one another within patients cross-sectionally or longitudinally. Understanding these relationships among symptoms may guide more precise symptom measurement in ways that increase the likelihood that patients with certain nasal and sinus symptoms also have objective evidence of opacification.

We used exploratory factor analysis (EFA) to assess the latent structure of nasal, sinus and other common, relevant symptoms at cross-section for three separate time points, as well as the change in these symptoms over time. By latent structure of symptoms, we mean the otherwise unseen patient attributes driving the manifestation of symptoms. While prior studies have used EFA applied to CRS symptoms at one point in time, they have utilized the Sino-nasal Outcome Test (SNOT) family of questionnaires, designed to assess treatment effectiveness among patients known to have CRS (Browne et al., 2007; Claire Hopkins, Browne, Slack, Lund, & Brown, 2007). SNOT assesses symptom severity in a two-week recall window, so cannot be used to evaluate compliance with EPOS duration criteria, and does not evaluate symptom frequency (Claire Hopkins et al., 2007; Wj et al., 2012). The questionnaire utilized in this study incorporated questions assessing frequency of EPOS defined symptoms, as well as frequency of

severe and related symptoms, in order to assess a broad range of potentially CRS-associated manifestations. Understanding how symptoms may group at one point in time and change over time could allow development of more precise approaches to symptom measurement; and also allow development of different biologic rationales for how these symptoms may group the way they do.

Methods

Study population and design

A total of 200,769 Geisinger Clinic primary care patients over the age of 18 years with both electronic health record (EHR) and race/ethnicity data were eligible for participation in this study. From these patients 23,700 were chosen to be recipients of a series of questionnaires utilizing a sampling scheme that has been previously descried (Hirsch et al., 2017; Tustin et al., 2017). In brief, using a stratified random sampling method to over-sample both racial/ethnic minorities as well as those with higher likelihoods of CRS using *International Classification of Diseases* (ICD-9) and *Current Procedural Terminology* codes in EHR data, patients were selected to receive self-administered questionnaires through the mail (Hirsch et al., 2017; Tustin et al., 2017).

Participants who returned the baseline questionnaire (n = 7847) were followed for 16 months, from April 2014 to August 2015, and received two additional questionnaires at six months and 16 months. Non-responders were sent questionnaires one or two additional times. The questionnaires were diverse in terms of information requested, providing information

about a spectrum of symptoms including presence, frequency, severity, and bother of a range of symptoms associated with CRS and co-morbid conditions like headache disorders and asthma (**Table** 1, Hirsch et al., 2017; Tustin et al., 2017). Each questionnaire included 37 common questions, each with the same response options (how often the symptom occurred in the past three months as 1 = never, 2 = once in a while, 3 = some of the time, 4 = most of the time, or 5 = all the time; **Table 2**). A total of 21 questions were about the presence, severity, and degree of bother of CRS nasal and sinus symptoms were incorporated while the remaining questions assessed presence of four asthma symptoms; four allergy symptoms; three ear symptoms; and five constitutional and other related symptoms (**Table 2**).

Data collection

The baseline questionnaire was mailed in April 2014, the 6-month follow-up in October 2014, and the 16-month follow-up in August 2015. These consisted of 94, 87, and 79 questions respectively, but the current analysis focused on the 37 questions that were common to all three (**Table 2**). After questionnaires were received, each was scanned and then data was double-checked and verified. A total of 7834 persons returned the baseline questionnaire (responding to at least one of the 37 questions of interest), 4945 returned the 6-month follow-up questionnaire, and 4584 returned the 16-month follow-up questionnaire.

Skip patterns were present in the questionnaires, by which patients would be asked to skip blocks of questions if the responses to these questions could be completely determined by previous responses. This occurred only when patients indicated that they had not experienced the symptom(s) of interest, making further questions pertaining to that symptom irrelevant.

These skip patterns were accounted for by filling in implied responses when skip-pattern missingness was present.

Analytic variables

The European Position Paper on Rhinosinusitis and Nasal Polyps subjective (EPOSs) criteria were used to classify patients as current, previous, or never CRS based on patient reported symptoms from only the baseline questionnaire. EPOSs criteria require three months of obstruction or anterior or posterior discharge with one other of the cardinal symptoms of smell loss, facial pain, or facial pressure, lasting three or more months. Patients were classified using questionnaire responses, specifically lifetime and previous 3 months of symptoms, being labeled as current CRS, if they met EPOSs CRS criteria in the three months before the baseline questionnaire; as past CRS if they met these criteria in their lifetime but not in the three months before the baseline questionnaire; and never CRS if they never met these criteria in their lifetime. The questionnaire has been previously described, from the Chronic Rhinosinusitis Integrative Studies Program (Hirsch et al., 2017; Wj et al., 2012), and included income and education information at baseline. Other demographic characteristics including age, sex, and race/ethnicity, as well as health information such as body mass index (BMI, measured in $\mathrm{kg/m^2}$), were collected via electronic health record data.

Statistical Analysis

The goals of the analysis were to identify the underlying structure, if present, among the 37 survey questions at each questionnaire time point, and then among the change in these

symptoms over time, from baseline to 6-month follow-up and from 6-month follow-up to 16-month follow-up questionnaires. Of the 7847 patients who returned the baseline questionnaire, the analysis included the 3535 patients who returned all three questionnaires with no more than 5 missing values for the 37 questions for any single questionnaire. We did not want to impute values for subjects with many missing questions since the primary goal of the analysis was to evaluate the underlying latent structure of the patterns of symptom reporting, and a large portion of patients were only missing a small number of responses. For missingness for subjects with five or fewer missing values, which we assumed to be at random, multivariate imputation by chained equations was conducted to impute missing values for patient questionnaires that were included in this study (3.5%), utilizing only information within each survey. This imputation was carried out via the mice R package using the predictive mean matching method (Buuren & Groothuis-Oudshoorn, 2011). Once data were finalized for each patient questionnaire, two change scores were calculated as the difference between each person's adjacent questionnaires (baseline to 6-month and 6-month to 16-month).

Due to the exclusion criteria utilized in this study, not all patients were included in the final analysis. Summary statistics of demographic, health, and socioeconomic information was computed and compared between the included individuals in this analysis and those who were excluded. In addition, lasagna plots were examined in order to visually assess the transitions between individual question responses over the 3 questionnaire duration of the study period (Figure 1, Swihart et al., 2010).

Exploratory factor analysis was utilized as there were multiple hypotheses and little *a priori* knowledge of the underlying structure of symptom reporting. Recognizing the ordinal

scaling of the data, implied Pearson (polychoric) correlations were estimated among the 37 questions for the three cross-sectional questionnaires, using the quick two step procedure as implemented by the psych R package (Revelle, 2017). These correlations were then utilized in exploratory factor analyses with ordinal variables. Meanwhile, Pearson correlation matrices were calculated for each of the two change scores as the difference score distribution appeared symmetric and contained more levels than practical for polychoric correlations.

For each of the five EFAs (three cross-sectional and two differences), a factor analysis was conducted fitting loadings estimates and communalities applying the ordinary (unweighted) least squares (OLS/ULS) procedure to correlations estimated as just described. An oblimin rotation for each factor analysis was utilized in order to allow for correlations among factors (Revelle, 2017). In EFA settings, determining the number of factors is a key step in identifying factor structure. Commonly, many methods are utilized in order to assess which selection is most appropriate, each with different optimality criteria driving different interpretations of results. In this study, biological interpretability and parsimony were stressed, in accordance with analyses and considerations provided in the previous chapter, the qualitative method of examining scree plots and the quantitative parallel analysis method were taken together to determining the optimal number of factors to extract. The scree plot displays eigenvalues of the correlation matrix in rank order by size from largest to smallest (x-axis = size rank, y-axis = eigenvalues) to assess the location of a clear “elbow” shape where the slope of the curve changed from rapid decline in eigenvalues with increasing rank to a flattening of the curve, as per Cattell’s Scree test (Cattell, 1966). Meanwhile, parallel analysis compares eigenvalues from random data matrices with uncorrelated item responses with observed

eigenvalues: the number of ranked observed eigenvalues greater than the randomly generated ones is the number of factors retained (Humphreys & Jr, 1975). Once factor loadings were extracted, factor scores were estimated for each identified factor for each patient using item response theory (IRT) based scores for polytomous items for each of the three surveys (Kamata & Bauer, 2008). These estimated factor scores were computed as a measure of the strength of each latent factor for each patient. These estimated factor scores were compared across EPOSs CRS status groups (current, previous, never). A multivariate analysis of variance (MANOVA) was fit in order to compare the mean multivariate factor score (vector of the estimated factor scores) between EPOSs CRS status groups for factor scores which appeared to follow an approximately normal distribution. One factor score had a mixed-scale distribution, with a considerable proportion of individuals having a low (at the IRT-lower bound) value and the remaining individuals distributed relatively continuously among higher values. To relate this score to EPOS status, a logistic regression was computed to estimate the odds of having a low factor score as a function of EPOSs CRS group, and a linear regression was used to estimate the mean factor scores by EPOSs CRS group for those patients who did not have the lower bound factor score.

We also sought to assess whether or not the baseline – 6 month difference captured more variability than the 6 month – 16 month difference. To this end, each difference EFA communality was extracted and then compared by time period (baseline-6 month versus 6 month-16 month), using a Wilcoxon signed rank test. We hypothesized higher mean communality values in the baseline – 6 month period EFA than the 6 month – 16 month period, corresponding to higher stability over a shorter period for change.

Sensitivity Analysis and Diagnostics

Diagnostics

Kaiser-Meyer-Olkin (KMO) factor adequacy was evaluated for each computed correlation matrix to further assess the appropriateness of factor analysis. KMO mean square error (MSA) statistics of 0.96, 0.96, 0.95 were observed for baseline, 6-month follow-up, and 16-month follow-up questionnaires, respectively. The two change score difference correlation matrices yielded KMO MSAs of 0.91 and 0.90 for the first and second differences, respectively. These KMO statistics, all of which were greater than 0.9, indicated a very high degree of common variance, and supported a conclusion that our covariance matrices were very well suited to be subjected to factor analysis.

Sensitivity to factoring method

Each factor analysis was refit using weighted least squares, principal factors, maximum likelihood, and generalized least squares to ensure the qualitative interpretation of the loadings was not conditional on the factoring method. We selected ordinary least squares as the final factoring method as OLS produces unbiased rotated factor loadings, and has desirable characterizes at large sample sizes (Lee, Zhang, & Edwards, 2012). Loading matrices, communalities, and inter-factor correlation matrices were examined. Loadings matrices, which may contain entries ranging from -1 to 1, provide a measure of the strength of the relationship between each question and each of the extracted factors, while the commonalities for each

question, which range from 0 to 1, are interpreted as the fraction of how much each question's variability is explained by the utilized factor model. Finally, inter-factor correlation matrices examined each factor's relations with the other factors that were derived from the final EFA.

Imputation

To evaluate the sensitivity of results to missing data and imputation, a total of 100 imputed datasets were generated from the original dataset with missingness using the same multiple imputation methodology (mice) as previously stated. Latent continuous (polychoric) correlations were calculated and compared across imputed datasets for each questionnaire item. Each of these 666 (all of the bivariate correlations among the responses to the 37 questions) pairwise correlations' standard deviations were computed using the 100 imputed datasets, and were examined. Across these pairwise correlations, 99.5% of standard deviations were below 0.0064, 0.0089, and 0.0036 for the baseline, 6-month follow-up, and 16-month follow-up questionnaires, respectively, suggesting that the impact of random imputation on correlation matrices and subsequent factor analyses was minimal.

Results

Description of study subjects

The 3535 patients included in the analysis were first compared to the 4312 respondents of the baseline questionnaire who were not included (**Table 1**). The two groups were similar on sex distribution (37.8% vs. 36.9% male, respectively) and mean body mass index (BMI, 30.0 vs. 30.3 kg/m^2, respectively). However, included and excluded patients differed on a number of

other study variables, including age (57.5 vs. 53.2 years on average, respectively), race/ethnicity (94.0% vs. 87.5% white, respectively), and socioeconomic status (32.9% vs. 25.1% earned over $50,000 annually, respectively).

It was observed that across time, individuals experienced varying degrees of changing symptoms, as exemplified by responses to the question (number 3) about the frequency of post nasal drip across visits (**Figure 1**). Although symptoms at baseline generally predicted symptoms over time, it was common for symptoms to change by one frequency category, and some patients changed by two or more. Those who answered "never" having post nasal drip in the previous 3 months at baseline typically responded having low frequency (never or once in a while) of the symptom at 6 months and 16 months (**Figure 1**). Similarly, those who responded experiencing the symptom "all of the time" at baseline, were more likely to experience the symptom often at 6 months and 16 months. This pattern was evident in other questions as well (results not shown); data are displayed for question 3 because it had a relatively uniform distribution of responses at baseline (the other questions had larger proportions of subjects who reported never experiencing the symptom).

Cross-sectional EFAs

For each of the three cross-sectional EFAs, scree plot results supported the extraction of five factors (see appendix). Parallel analysis suggested the retention of five factors for baseline and 6 month questionnaires, and six factors for the 16-month follow up. For comparability, 5 factors were extracted from each of these questionnaires. Each of the structures in these three EFAs was similar, and the content of the five factors was the same for each, with one factor

each for congestion and discharge symptoms, pain and pressure symptoms (including headache), asthma and constitutional symptoms, ear and eye symptoms, and smell loss (**Table 3** for baseline EFA, other two cross-sectional EFAs, see appendix). Factor loadings, or the degree to which any specific question was related to a specific latent factor, were consistent across all three questionnaires (similar to results in **Table 3** for baseline, other cross-sectional EFA results not shown). Most observed communalities were high, indicating that the factor models well-represented these questions which were included in the analysis (**Table 3**). A few low communalities were observed for bad breath (0.26), fever (0.34), cold/flu symptoms (0.37) and fatigue (0.39), suggesting that our five-factor model did not account for much of the variability in these symptoms, and thus they did not load heavily on any single factor. After using the baseline model to estimate factors within individuals at baseline, the inter-factor correlations resulting from oblimin rotation ranged from 0.30 to 0.64 (**Figure 2**).

Longitudinal difference EFAs

Analysis results also supported five-factor models for each of the two longitudinal difference EFAs. Symptoms identified to load on single factors in the difference analyses indicates that these symptoms change together and in the same direction over time. Notably, the two difference EFAs (both for 6 and 10 month durations) yielded nearly identical factors (**Table 4** for change from 6-month to 16-month questionnaires, see appendix) which displayed a fairly similar structure to the factors identified in each of the cross-sectional EFAs (**Table 3**). In order to compare model fit between the two difference EFAs, a Wilcoxon signed rank test was utilized to test the hypothesis that the baseline to 6-month difference EFA explained more

variability than the 6 to 16-month difference EFA by examining the difference in individual variable commonalities between the two EFAs. The baseline to 6 month difference EFA had a significantly greater average communalities than the 6 month to 16 month difference EFA (p-value = 0.002; see appendix).

Factor Scores

Utilizing the factors from the baseline questionnaire EFA, factor scores were estimated and compared between EPOSs CRS groups (current, previous, and never at baseline; **Figure 4** for factor 1, results for other factors not shown). A MANOVA was fit, comparing the four factor scores in the three CRS status groups simultaneously (omitting the 4^{th} factor, smell loss, as it appeared to be non-normally distributed), and this indicated a significant difference between mean factor scores between groups (p-value < 0.001; **Figure 3**). We observed factor scores were higher, in descending order, for current, past, then never. We also observed that CRS Factor scores showed a weaker, but similar structure as individual questionnaire responses, with low factor score values correlating more strongly with low scores or response in the following survey and vice versa (**Figure 2, 4**). Utilizing the factor scores from the 4^{th} factor (smell loss), two regressions were conducted, a logistic regression estimating the odds of having a lower bounded factor score (-4) as a function of EPOSs CRS status, and a linear regression comparing average factor scores for those with factor scores above -4 as a function of EPOSs CRS status. Those with EPOSs CRS current at baseline had an odds ratio of experiencing lower bound factor scores of 0.05 (95% CI: 0.04, 0.06) compared with EPOSs never, and EPOSs previous had an odds ratio of experiencing lower bounded factor 4 factor scores of 0.15 (95%

CI: 0.12, 0.18) compared with the EPOSs never group. For those above the lower bounded factor score, a linear regression was conducted. Those at EPOSs CRS current had a factor 4 score 0.48 higher than those at EPOSs CRS never at baseline (95% CI: 0.35, 0.60) while those at EPOSs CRS previous had a factor 4 factor score of 0.22 higher than those at EPOSs CRS never at baseline (95% CI: 0.1, 0.34).

Discussion

Exploratory factor analysis was conducted as a measurement exercise to better understand the relationship and categorization of nasal and sinus, asthma, headache, constitutional, allergy, and ear symptoms utilizing both cross sectional symptom questionnaire responses and changes in symptom responses over time. All five EFAs presented consistent findings of five underlying factors that were identifiable as congestion and discharge, pain and pressure, asthma and constitutional, ear and eye, and smell loss factors. The baseline EFA was used to estimate five factor scores within subjects, and all five were higher in subjects who met EPOSs current CRS criteria and lowest in those who met EPOSs never CRS criteria. The 37 questions utilized in this analysis were developed to capture a wide range of overlapping symptoms that occur in several co-morbid conditions; and to evaluate both the frequency and severity of symptoms. Understanding how symptoms cluster within visit and across visits can provide useful information that can aid clinical practice, inform symptom measurement in CRS patients, and lead to hypotheses about the pathobiology underlying these symptoms.

We hypothesized several patterns of results in the EFAs. One the one hand, if there is an underlying construct of CRS that can be measured with six cardinal symptoms that are mainly

interchangeable, as the EPOS criteria suggest, then these cardinal CRS symptoms could be expected to load on a single factor. On the other hand, there are specific sinuses in which inflammation has been associated with specific symptoms, such as maxillary sinus inflammation associated with facial pain and pressure and ethmoid sinus inflammation associated with smell loss. This pathobiologic consideration would suggest that at least three factors would be identified with the cardinal CRS symptoms. We found that the 37 symptoms identified five factors, and the six cardinal CRS symptoms loaded on three factors, both in cross-sectional and longitudinal EFA models. This stability, and coherence to real biological processes may provide some evidence that these five factors may have an underlying common pathobiology.

Three of the five factors were composed of questions that are components of the EPOSs criteria for CRS, specifically the nasal congestion and discharge, facial pain and pressure, and smell loss factors, each of which is one of the cardinal EPOSs CRS symptoms (Wj et al., 2012). As these symptoms loaded on different factors, it is possible that the underlying pathobiology may be different from one another. While there was clear evidence that there was some longitudinal change in symptom reporting, large transitions (two or more steps on the Likert scale) were not very common. This suggests that most patients have long duration symptoms and perhaps a chronic pathobiologic process. We have previously reported that using the cardinal symptoms to define EPOSs CRS categories (current, past, and never) resulted in large transitions in patients meeting criteria for these categories over time; for example, among the subjects who met EPOSs current CRS at baseline, almost half did not meet criteria for current CRS six months later (Sundaresan, 2017, in Press).

In CRS, the sinuses are inflamed and swollen, and as such we can consider these factors in the context of paranasal sinus opacification. The maxillary and ethmoid sinuses, when congested or inflamed, can present symptoms of facial swelling and pain, which were present in the facial pain and pressure factor (Wald et al., 1981). Sphenoid opacification, although comparatively rare, can be associated with sometimes severe headaches (Sieskiewicz et al., 2011). Frontal and ethmoid sinusitis can be associated with smell loss and nasal discharge, consistent with the smell loss factor which was observed (Chang, Lee, Mo, Lee, & Kim, 2009). The sinus symptoms that tended to cluster in our five factors have generally strong sinus opacification correlates (Chang et al., 2009; Sieskiewicz et al., 2011; Wald et al., 1981). It is possible that the factors that were identified by the EFA procedure may be associated with sinus opacification, an analysis that is currently underway. While the factor analysis was performed in over 3500 subjects, sinus CT scans were obtained from 646 of these subjects. Furthermore, the ear and eye symptoms seen to predominate in factor 5 may be measuring allergy presence and severity.

Because of the factor rotation method chosen, non-orthogonality between factors was allowed as a means to separate symptoms into distinct factors insofar as possible. Thus, examining the correlation between factors may provide some insight into the symptom relationships. The congestion and discharge factor was moderately correlated with both the pain and pressure factor ($\rho = 0.67$) as well as the ear and eye symptom factor ($\rho = 0.63$, **Figure 1**). These correlations are not entirely unexpected as there are several pathobiologies that could drive these patterns; other drivers connecting reporting of different symptoms could also be at work.

Most questionnaire responses were well represented by the observed factor model as measured by the communalities, which give a quantitative measure of the variability of each question explained by our final five-factor model. In the EFA of the baseline questionnaire, most of the EPOSs core questions had commonality values above 0.6, indicating that the model accounted for a majority of the variance in the reporting of these symptoms. In contrast, several symptoms did not load heavily onto any factors in the baseline EFA, and as such also had the lowest commonalities of 0.26 (bad breath), 0.37 (cold symptoms), and 0.39 (fatigue), suggesting that the model failed to capture much of the variability in those questions. In an EFA setting, we do not necessarily expect all communalities to be high. Instead, these low communality values reveal that either the drivers of these variables were different than the five observed factors (high unique variance) or that these symptoms were subject to higher measurement error than others.

It is common, in social and medical sciences, for factor analyses to include data from a single point in time or in a context where time is unimportant (Browne et al., 2007). In this study, we were able to incorporate repeated observations, providing us with not only three responses for each symptom question, but also explicit measures of how symptoms changed over time. EFA theory hypothesizes that there are real underlying mechanisms, including common pathobiology, reporting phenomena, or other reasons which manifested itself in the clustering of symptoms into the observed factors. If this hypothesis was correct, we would expect factor composition (loadings) to be invariant to time (i.e. no seasonality); and if there was sufficient variation in symptom reporting over time, we would expect to not only see factors present themselves across time, but we would expect the changes in symptoms to do so

according to these same factors. Without sufficient symptom changing, we would likely not have strong enough differences to identify these same factors. In this study, we did observe the same factors in EFAs of cross-sectional responses as well as in the differences in reporting over time, a finding consistent with the idea that these are real constructs driving the observed symptoms. In the difference EFAs, there were almost always lower commonalities compared with the cross sectional EFAs. Because responses to questionnaire items over time did not evidence large changes (i.e., most change scores fell between -1 and +1, and many at 0), we would expect the commonalities in differences to be smaller than for their cross-sectional counterparts. These differences in communalities also could be due to the different correlations utilized, polychoric (implied Pearson correlations) versus Pearson correlations. In addition, measurement and reporting error were likely larger in the difference measures as we were combining together potentially two (not necessarily independent) error terms.

The communalities in the difference scores were considerably lower than those in the cross sectional EFAs ranging from 0.09 (fever) to 0.75 (smell loss) in the change from 6-month to 16-month EFA and from 0.26 (bad breath) to 0.95 (smell loss) in the baseline EFA. These commonalities show that smell loss was a very persistent symptom. The time duration in the two change EFAs were not the same; the first change measure was over six months and the second was over 10 months. We would expect commonalities to be lower for the longer duration EFA, and we found this to be the case. The mean communality of the first change measure (0.392) was significantly larger than for the second (0.356, p-value = 0.001), which was in line with expectations, suggesting that the difference from baseline to 6 months (6 month

duration) captured variability better than the difference from 6 month to 16 month questionnaires (10 month duration).

The multidimensional mean factor scores were compared between the three EPOSs CRS groups (current, past, never) using MANOVA, logistic regression, and linear regression: a significant difference was found, indicating a difference in factor score distributions between the EPOSs CRS groups (P-value < 0.001). For all five factors, the mean estimated factor scores tended to be highest among current CRS subjects, next highest for past CRS, and lowest for never CRS. This result is not in itself surprising as the CRS groups here were determined by the EPOSs definition of the disease, which itself is based on many of the symptoms in the factors. However, the EFA included many questions beyond those used to define EPOSs CRS status. The higher factor scores comprised of eye, ear, asthma, constitutional, and headache symptoms may represent the common co-occurrence of allergy, asthma, and headache disorders, for example, among CRS patients.

While there has been some prior work on CRS factors at a single point in time with the SNOT-20 and SNOT-22 questionnaires, there has been no prior work on CRS factors using longitudinal information on symptoms (Browne et al., 2007). Previous studies have examined the decomposition of CRS and related symptoms using the Sino-nasal Outcome Test (SNOT)-20 and (SNOT)-22 questionnaire which measures “symptoms and social/emotional consequences of rhinosinusitis” through a range of symptom and health-related quality of life questions in 20 or 22 Likert-scale questions (DeConde, Bodner, Mace, & Smith, 2014; C. Hopkins et al., 2006). These questions ask the participants to consider physical, functional, and emotional symptoms they have experienced in the previous 2-week period (DeConde et al., 2014; C. Hopkins et al.,

2006). SNOT was designed to provide a single measure of patient quality of life and CRS-related symptom severity, implicitly suggesting that each question provides information regarding a single CRS construct or factor (Browne et al., 2007). One study found questions from the SNOT-22 decomposed into five clear rhinologic symptoms, extranasal rhinologic symptoms, ear & facial symptoms, psychological dysfunction, and sleep dysfunction factors, not a single factor as the mission of the SNOT surveys may suggest (DeConde et al., 2014). Similarly, an analysis of SNOT-20 revealed four underlying latent factors, rhinological symptoms, ear and facial symptoms, sleep function, and psychological function (Browne et al., 2007). Taken together, both of these studies suggested that question sets typically thought of measuring only CRS symptom severity or CRS-related quality of life, were actually measuring a variety of unobserved dimensions. Although the set of 37 questions utilized in our study is much different in scope and aim than the SNOT questionnaires, the results were consistent in revealing several factors.

Limitations and Further Work:

We observed both similarities and differences between patient characteristics in the subject who completed the baseline questionnaire who were included and excluded in the EFAs. Subjects in the EFA analysis, who returned all three questionnaires without excessive missing data, were more likely to be white, more highly educated, and with higher incomes. This may have resulted in selection bias that could have influenced the results. This project relied extensively on questionnaire question responses. While direct and easy to interpret or compare, survey methodologies similar to this encourage respondents to only report questions

that were asked about, potentially missing symptom associations and relationships with latent factors. In addition, the is the potential of same source bias impacting results by which some individuals report in a systemic manner not necessarily associated with symptoms, such as some individuals always or never reporting experiencing symptoms. Finally, while we found strong evidence of clustering among 37 symptoms within visits and over time, the ultimate utility of the findings will be in comparison to sinus opacification, which awaits further analysis in a subsect of the included subjects.

This EFA generated several hypotheses, mainly that the underlying factors identified by the procedure are measuring real biological phenomena, including distinct sinus opacification, allergies, and asthma severity. While interesting, studies examining objective measures of the presence of these conditions along with these symptom questions are needed to provide substantive evidence of this relationship.

Conclusion

In an analysis of 37 nasal and sinus, allergy, ear, asthma, headache, and constitutional symptoms, we identified five underlying factors – congestion and discharge, pain and pressure, asthma and constitutional, ear and eye, and smell loss – that were consistent in three cross-sectional and two longitudinal change EFAs. Questions assessed presence, severity, bother, and frequency of all 37 symptoms. The models generally explained a large proportion of the variation in these symptoms within visits, and symptoms like smell loss showed much persistence across visits. The findings have implications for how to identify patients with CRS using questionnaires and may suggest significant misclassification in EPOS approaches to CRS

identification. They may explain why patients who meet EPOS criteria for CRS often do not have evidence of sinus opacification. More direct evidence awaits analysis of the sinus CT data in a subset of these subjects.

Tables & Figures

Table 1. Demographic information of the 3535 patients included in the current analysis and the 4312 patients who returned the baseline questionnaire but were not included in the current analysis.

	Excluded	**Included**
Male, n (%)	1591 (36.9)	1335 (37.8)
Age, years, mean (SD)	53.2 (16.8)	57.5 (14.8)
Smoking status		
Never, n (%)	2253 (52.2)	2053 (58.1)
Former, n (%)	1299 (30.1)	1100 (31.1)
Current, n (%)	760 (17.6)	382 (10.8)
Income:		
< $25,000, n (%)	1599 (37.1)	1021 (28.9)
$25,000-$50,000, n (%)	1098 (25.5)	970 (27.4)
> $50,000, n (%)	1083 (25.1)	1163 (32.9)
Body mass index, kg/m^2, mean (SD)	30.3 (7.05)	30.0 (6.93)
Education level		
High school, n (%)	1608 (37.3)	1209 (34.2)
Some college, n (%)	1364 (31.6)	979 (27.7)
College graduate, n (%)	977 (22.7)	1171 (33.1)
Race/ethnicity		
White, n (%)	3372 (87.5)	3323 (94)
Black, n (%)	264 (6.1)	78 (2.2)
Hispanic, n (%)	276 (6.4)	134 (3.8)

Table 2. Questions for the three cross-sectional questionnaires. Question responses were on a 5-item Likert scale*

Item #	Question text
On average, how often in the past * months have you had …	
1	… blockage of your nasal passages (nasal congestion)?
2	… nasal discharge that was yellow or green in color?
3	… post-nasal drip?
4	… loss of sense of smell?
5	… facial pain?
6	… facial pressure?
Check the box that describes how often each problem has happened in the past † months, on average	
7	… both of my nasal passages have blockage
8	… at least one of my nasal passages is completely blocked
9	… I have been very bothered by, my blocked nasal passage(s)
10	… I have a lot of nasal discharge
11	… I have to blow my nose more than 10 times a day because of my nasal discharge
12	… I have been very bothered by my nasal discharge
13	… I have been coughing after I eat or lie down
14	… I have had mucus in my throat that felt like a lump or blockage
15	… I have been very bothered by my post-nasal drip
16	… I have not been able to smell anything
17	… I have been very bothered by my loss of sense of smell
18	… On a scale of 0 to 10, my facial pain has been at least a 5 (0 = no pain, 10 = worst pain)
19	… I have been very bothered by my facial pain
20	… My facial pressure has been severe
21	… I have been very bothered by my facial pressure
Check the box that describes how often, on average, you had the following in the past † months …	
22	… headaches
23	… fevers
24	… coughing
25	… bad breath
26	… fatigue
27	… nasal itching
28	… sneezing
29	… eye itching
30	… eye tearing
31	… ear fullness
32	… ear pain
33	… ear pressure
34	… wheezing (breathing with whistling sound in chest)
35	… chest tightness
36	… shortness of breath

37	… cold/flu symptoms

* 1 = Never, 2 = Once in a while, 3 = Some of the time, 4 = Most of the time and 5 = All the time.

† For the baseline and 16-month follow-up = 3 months, for the 6-month follow-up = 6 months.

Table 3. Factor loadings and symptom commonalties from the exploratory factor analysis (EFA) of the 37 presence, severity, and secondary CRS symptom at baseline. The EFA was fit using ordinary least squares and an oblimin rotation (number of patients = 3535). Loadings less than 0.3 were omitted for readability. Communalities represent the fraction of each symptom's variability that was captured by the utilized five factor model.

#	Item Label	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Communalities
1	Blockage	0.65					**0.80**
2	Discharge discolored	0.49					**0.61**
3	PND	0.84					**0.78**
4	Smell loss				0.95		**0.89**
5	Facial pain		0.83				**0.85**
6	Facial pressure		0.76				**0.87**
7	Blockage both sides	0.58					**0.72**
8	Blockage complete	0.55					**0.73**
9	Blockage bothered	0.61					**0.81**
10	Discharge a lot	0.86					**0.84**
11	Blow nose 10x daily	0.82					**0.76**
12	Discharge bothered	0.84					**0.84**
13	Cough lie down	0.72					**0.73**
14	Lump in throat	0.69					**0.73**
15	PND bothered	0.84					**0.85**
16	Smell loss complete				0.97		**0.95**
17	Smell loss bothered				0.92		**0.91**
18	Facial pain 5+		0.83				**0.90**
19	Facial pain bothered		0.84				**0.91**
20	Facial pressure severe		0.77				**0.86**
21	Facial pressure bothered		0.78				**0.90**
22	Headaches		0.67				**0.48**
23	Fever			0.43			**0.34**
24	Coughing	0.46		0.5			**0.53**
25	Bad breath						**0.26**
26	Fatigue						**0.39**
27	Nasal itching					0.56	**0.53**
28	Sneezing	0.31				0.54	**0.51**
29	Eye itching					0.72	**0.62**
30	Eye tearing					0.6	**0.49**
31	Ear fullness		0.35			0.54	**0.62**
32	Ear pain		0.51			0.49	**0.65**
33	Ear pressure		0.47			0.46	**0.63**
34	Wheezing			0.8			**0.66**
35	Chest tightness			0.85			**0.78**
36	Shortness of breath			0.82			**0.68**

37	**Cold/flu symptoms**	**0.44**	**0.37**

Table 4. Factor loadings and symptom commonalties from the exploratory factor analysis (EFA) of the 37 presence, severity, and secondary CRS symptom changes from 6 to 16 months. EFA was fit using ordinary least squares and an oblimin rotation (number of patients = 3535). Loadings less than 0.3 were omitted for readability. Communalities represent the fraction of each symptom's variability that was captured by the utilized five factor model.

#	Item Label	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Communalities
1	Blockage	0.46					0.3
2	Discharge discolored	0.32					0.19
3	PND	0.49					0.28
4	Smell loss					0.68	0.47
5	Facial pain		0.66				0.47
6	Facial pressure		0.59				0.41
7	Blockage both sides	0.43					0.28
8	Blockage complete	0.34					0.23
9	Blockage bothered	0.52					0.4
10	Discharge a lot	0.72					0.5
11	Blow nose 10x daily	0.66					0.43
12	Discharge bothered	0.75					0.54
13	Cough lie down	0.34			0.33		0.29
14	Lump in throat	0.36					0.27
15	PND bothered	0.57					0.4
16	Smell loss complete					0.84	0.69
17	Smell loss bothered					0.68	0.48
18	Facial pain 5+		0.78				0.6
19	Facial pain bothered		0.79				0.63
20	Facial pressure severe		0.65				0.44
21	Facial pressure bothered		0.72				0.54
22	Headaches						0.14
23	Fever						0.1
24	Coughing				0.42		0.29
25	Bad breath						0.12
26	Fatigue						0.14
27	Nasal itching			0.35			0.19
28	Sneezing			0.38			0.25
29	Eye itching			0.5			0.29
30	Eye tearing			0.51			0.3
31	Ear fullness			0.64			0.41
32	Ear pain			0.53			0.33
33	Ear pressure			0.63			0.39
34	Wheezing				0.58		0.33
35	Chest tightness				0.64		0.4

36	Shortness of breath	0.6	0.37
37	Cold/flu symptoms	0.35	0.24

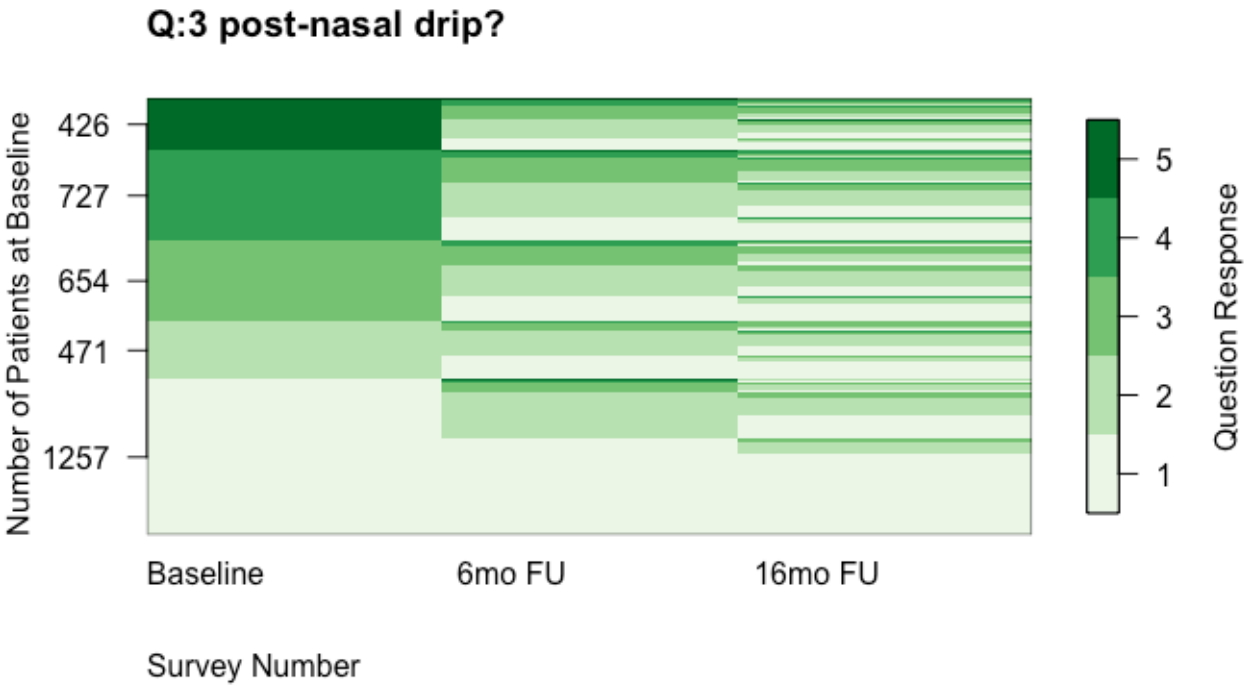

Figure 1. Lasagna plot displaying the proportion of individuals with each given response to the question "On average, how often in the past 3 months have you had post-nasal drip?" at baseline and 6 months and 16 months later (1 = Never, 2 = Once in a while, 3 = Some of the time, 4 = Most of the time, 5 = All the time). Y-axis values indicate the number of patients with each particular response at baseline.

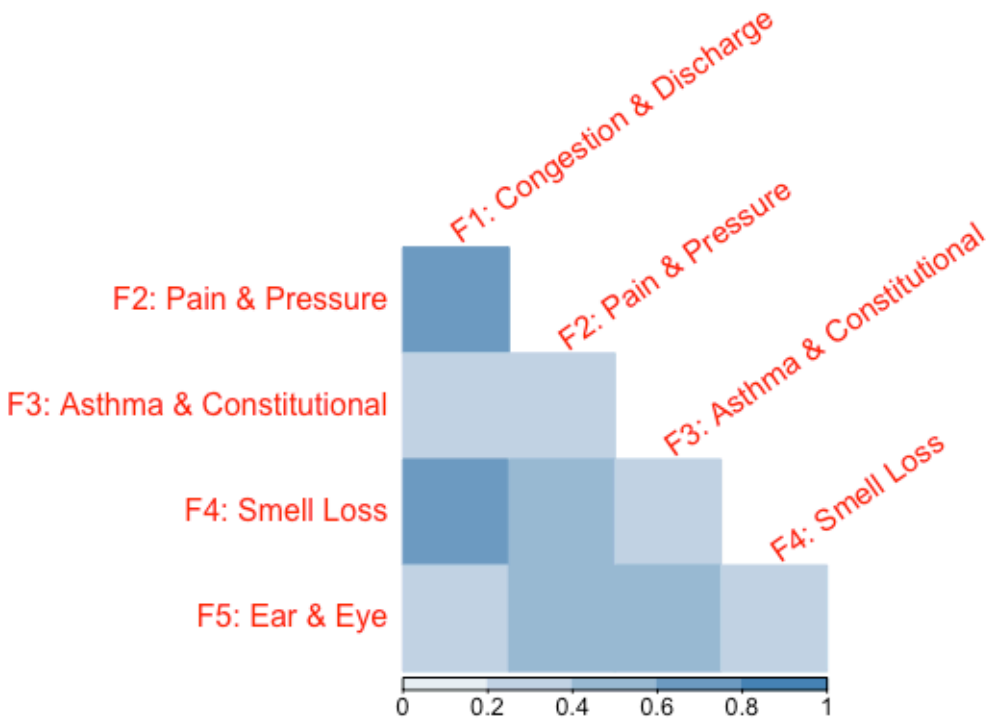

Figure 2. Inter-factor correlations at baseline from the baseline questionnaire exploratory factor analysis fit via ordinary least squares and an oblimin rotation.

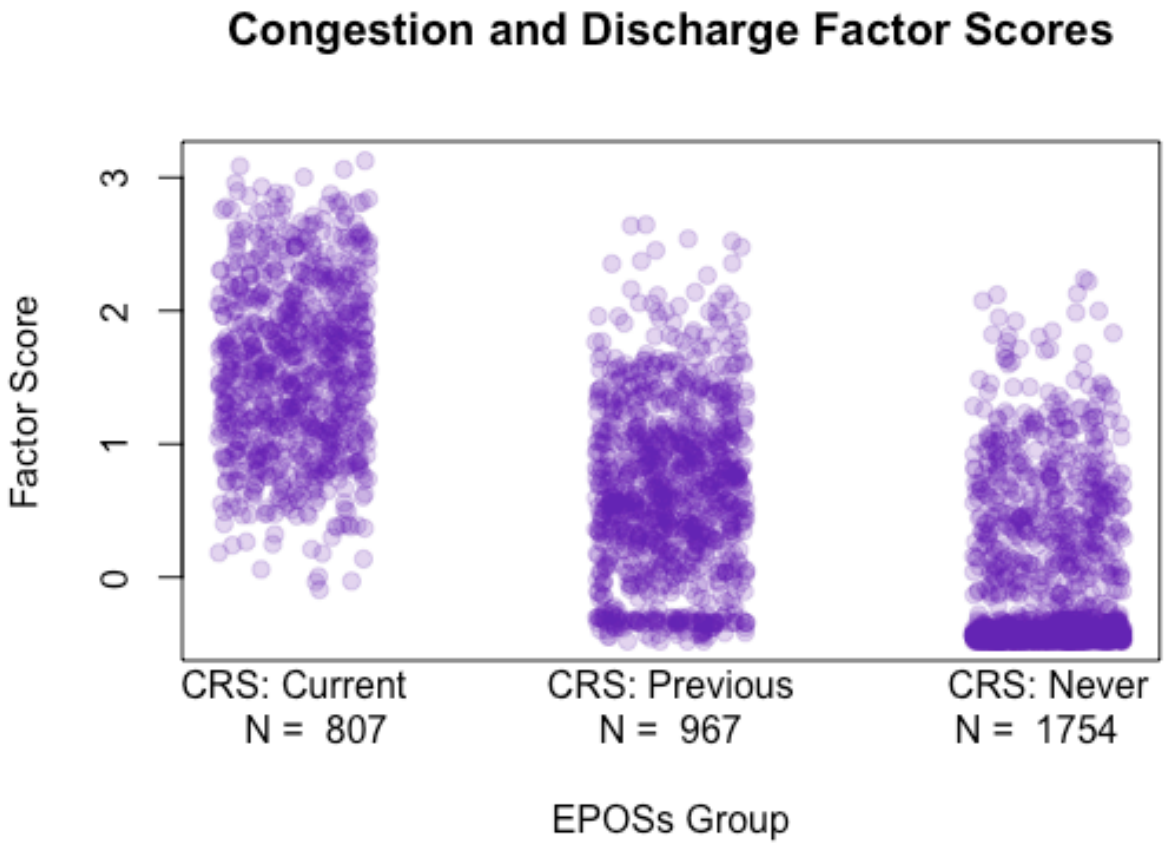

Figure 3. Factor 1 (congestion and discharge) scores by CRS EPOS groups at baseline. Factor scores across factors and CRS EPOSs groups (current CRS, previous CRS, never CRS) for the congestion and discharge factor at baseline with number of individuals (N) in each group. Factor scores were estimated by the Item Response Theory (IRT) based scores method. X-axis was jittered to improve readability.

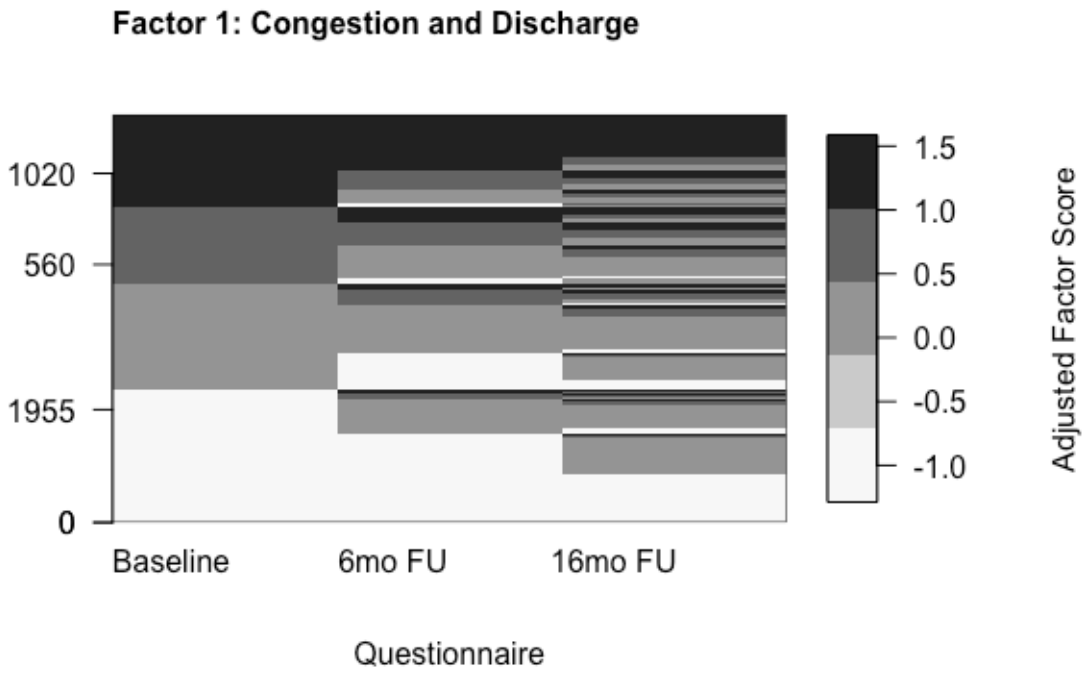

Figure 4. Continuous factor scores categorized to show longitudinal change across questionnaires for factor 1 (congestion and discharge). Factor scores were categorized as: factor score < -1 were assigned values of -2; between -0.5 and -1, assigned -1; between -0.5 and 0.5, assigned 0; between 0.5 and 1, assigned 1; and > 1, assigned 2. Y-axis labels indicate the number of patients at baseline in each adjusted factor score group. Factor scores were estimated by the IRT method.

Appendix

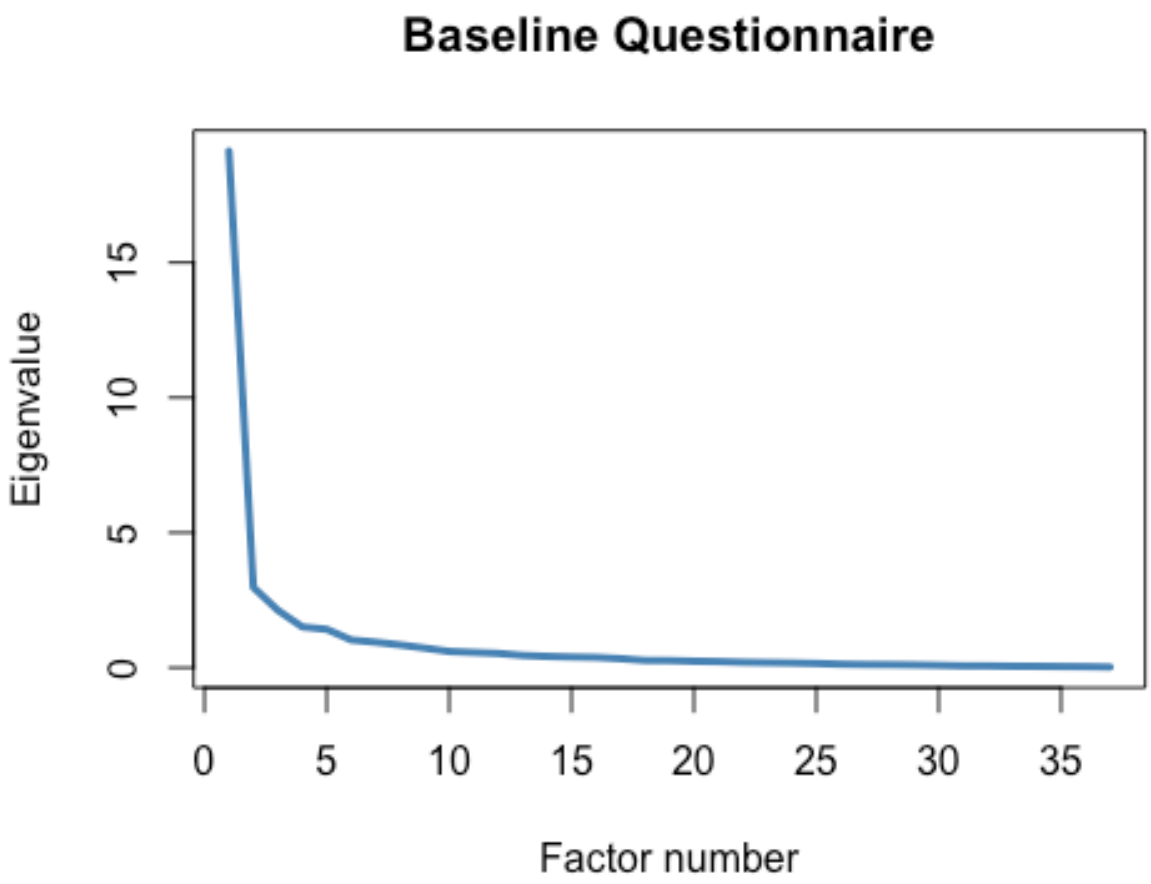

Figure A1. Scree plot for the baseline questionnaire displaying eigenvalues on the y-axis and their corresponding factor number on the x-axis.

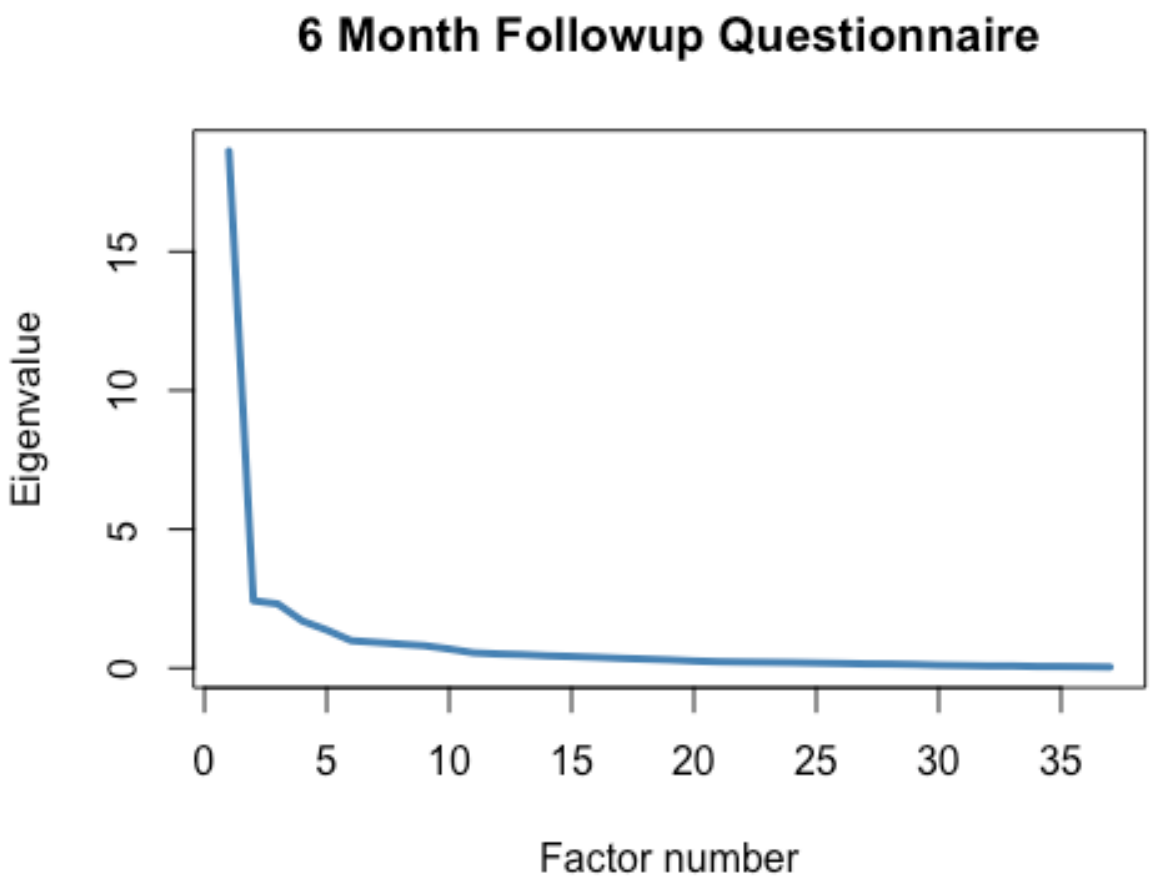

Figure A2. Scree plot for the 6 month follow up questionnaire displaying eigenvalues on the y-axis and their corresponding factor number on the x-axis.

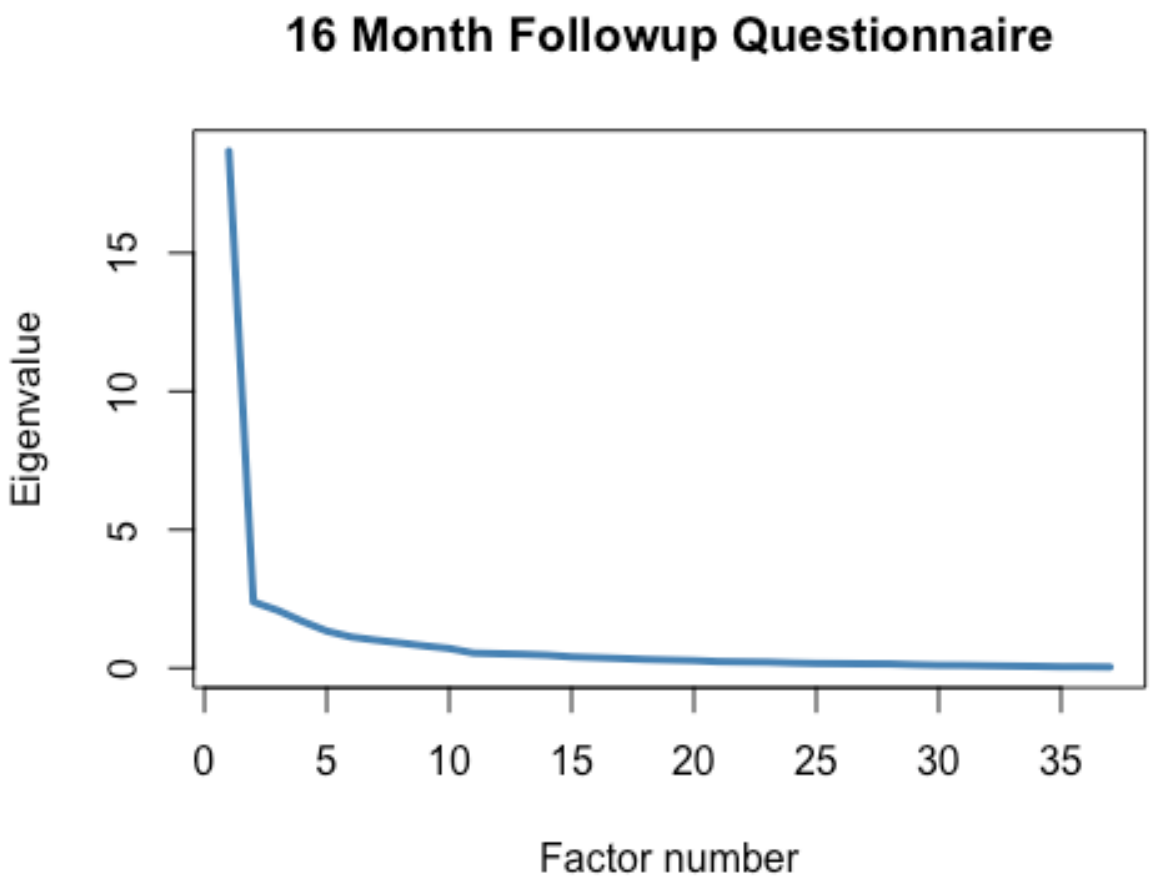

Figure A3. Scree plot for the 16 month follow up questionnaire displaying eigenvalues on the y-axis and their corresponding factor number on the x-axis.

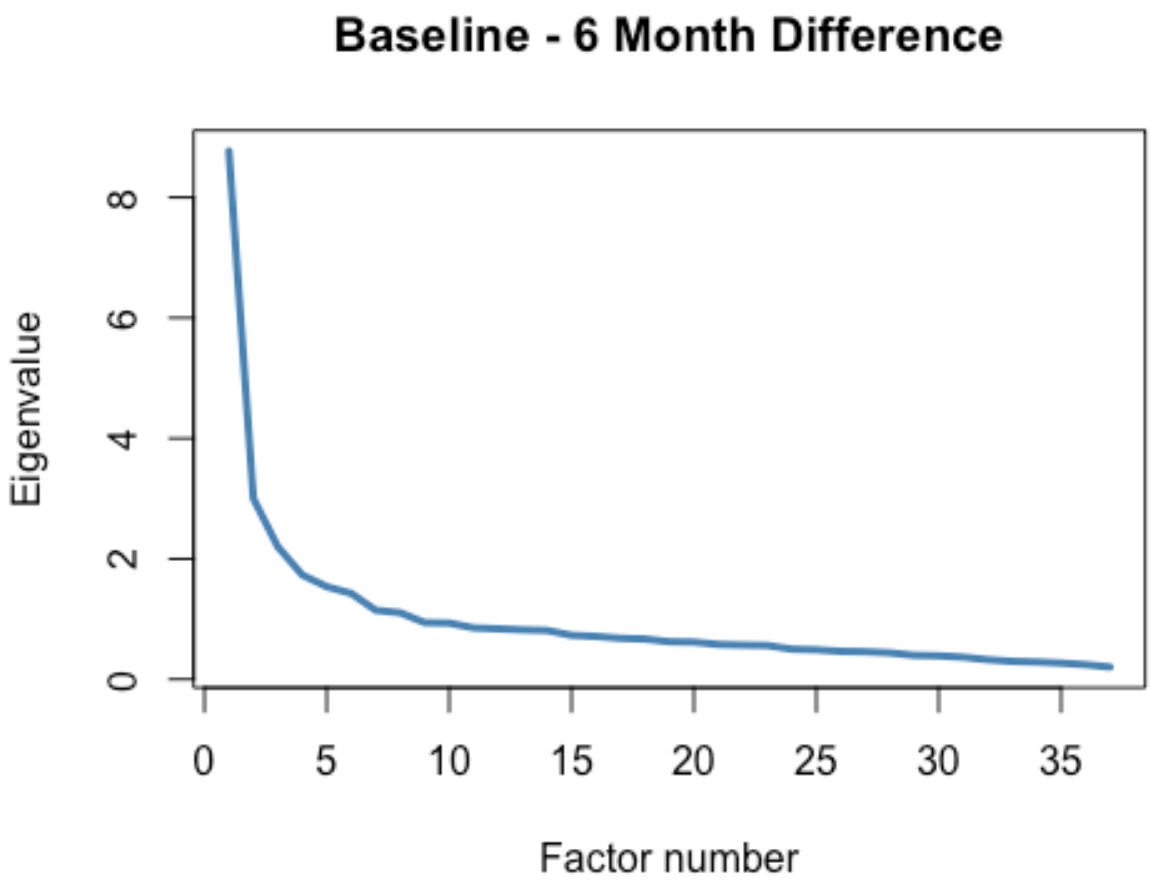

Figure A4. Scree plot for the first difference (baseline to 6 month questionnaires) displaying eigenvalues on the y-axis and their corresponding factor number on the x-axis.

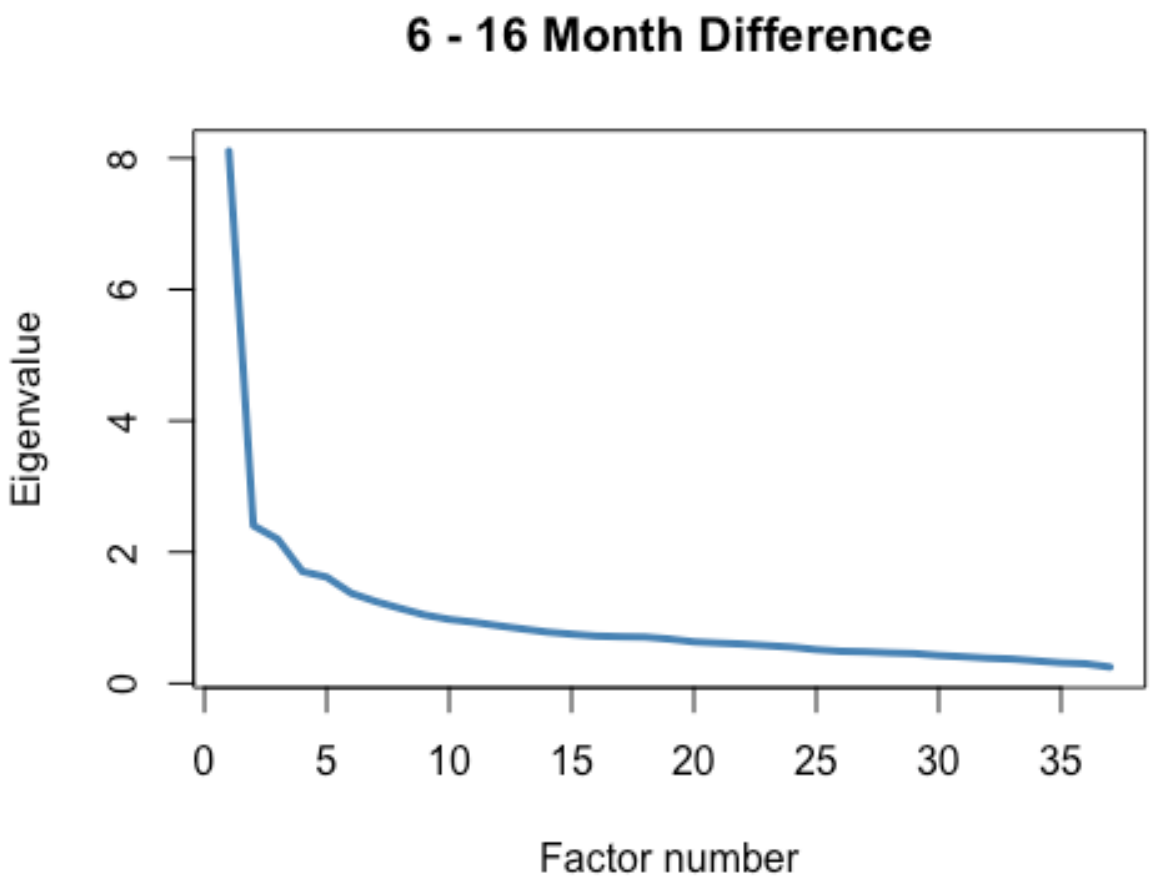

Figure A5. Scree plot for the second difference (6 to 16 month questionnaires) displaying eigenvalues on the y-axis and their corresponding factor number on the x-axis.

Table A1. Factor loadings and symptom commonalties from the exploratory factor analysis (EFA) of the 37 presence, severity, and secondary CRS symptom at 6 month follow up. The EFA was fit using ordinary least squares and an oblimin rotation (number of patients = 3535). Loadings less than 0.3 were omitted for readability. Communalities represent the fraction of each symptom's variability that was captured by the utilized five factor model.

#	Item Label	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Communalities
1	Blockage	0.51				0.65	0.65
2	Discharge discolored	0.33				0.46	0.46
3	PND	0.76				0.65	0.65
4	Smell loss				0.98	0.9	0.9
5	Facial pain		0.89			0.84	0.84
6	Facial pressure		0.86			0.85	0.85
7	Blockage both sides	0.5				0.61	0.61
8	Blockage complete	0.39				0.53	0.53
9	Blockage bothered	0.6				0.78	0.78
10	Discharge a lot	0.87				0.81	0.81
11	Blow nose 10x daily	0.82				0.7	0.7
12	Discharge bothered	0.85					0.82
13	Cough lie down	0.53		0.34			0.64
14	Lump in throat	0.55					0.64
15	PND bothered	0.79					0.78
16	Smell loss complete				0.97		0.94
17	Smell loss bothered				0.92		0.89
18	Facial pain 5+		0.87				0.88
19	Facial pain bothered		0.89				0.91
20	Facial pressure severe		0.83				0.84
21	Facial pressure bothered		0.85				0.88
22	Headaches		0.61				0.49
23	Fever			0.39			0.38
24	Coughing	0.4		0.5			0.57
25	Bad breath						0.33
26	Fatigue						0.43
27	Nasal itching					0.48	0.51

28	Sneezing	0.37			0.47	0.53
29	Eye itching				0.68	0.61
30	Eye tearing				0.6	0.5
31	Ear fullness				0.71	0.69
32	Ear pain		0.31		0.64	0.68
33	Ear pressure				0.66	0.69
34	Wheezing			0.81		0.68
35	Chest tightness			0.87		0.81
36	Shortness of breath			0.87		0.73
37	Cold/flu symptoms			0.45		0.46

Table A2. Factor loadings and symptom commonalties from the exploratory factor analysis (EFA) of the 37 presence, severity, and secondary CRS symptom at 16 month follow up. The EFA was fit using ordinary least squares and an oblimin rotation (number of patients = 3535). Loadings less than 0.3 were omitted for readability. Communalities represent the fraction of each symptom's variability that was captured by the utilized five factor model.

#	Item Label	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Communalities
1	Blockage	0.34	0.42				0.63
2	Discharge discolored		0.3				0.48
3	PND		0.68				0.64
4	Smell loss				0.98		0.9
5	Facial pain	0.88					0.86
6	Facial pressure	0.88					0.86
7	Blockage both sides	0.33	0.34				0.57
8	Blockage complete	0.34	0.38				0.57
9	Blockage bothered	0.4	0.43				0.72
10	Discharge a lot		0.84				0.78
11	Blow nose 10x daily		0.81				0.67
12	Discharge bothered		0.85				0.8
13	Cough lie down		0.38	0.42			0.57
14	Lump in throat		0.38				0.57
15	PND bothered		0.68				0.72
16	Smell loss complete				0.99		0.94
17	Smell loss bothered				0.93		0.9
18	Facial pain 5+	0.92					0.89
19	Facial pain bothered	0.92					0.89
20	Facial pressure severe	0.87					0.81
21	Facial pressure bothered	0.89					0.86
22	Headaches	0.58					0.49
23	Fever			0.35			0.35
24	Coughing		0.34	0.52			0.6
25	Bad breath						0.35
26	Fatigue			0.3			0.45
27	Nasal itching					0.46	0.56
28	Sneezing		0.42			0.44	0.56
29	Eye itching					0.63	0.61
30	Eye tearing					0.55	0.52
31	Ear fullness					0.59	0.68
32	Ear pain	0.41				0.5	0.68

33	Ear pressure	0.34		0.54	0.68
34	Wheezing		0.85		0.69
35	Chest tightness		0.85		0.75
36	Shortness of breath		0.89		0.72
37	Cold/flu symptoms		0.46		0.49

Table A3. Factor loadings and symptom commonalties from the exploratory factor analysis (EFA) of the 37 presence, severity, and secondary CRS symptom changes from baseline to 6 months. EFA was fit using ordinary least squares and an oblimin rotation (number of patients = 3535). Loadings less than 0.3 were omitted for readability. Communalities represent the fraction of each symptom's variability that was captured by the utilized five factor model.

#	Item Label	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Communalities
1	Blockage	0.68					0.49
2	Discharge discolored	0.43					0.24
3	PND	0.7					0.45
4	Smell loss				0.63		0.42
5	Facial pain		0.67				0.51
6	Facial pressure		0.61				0.51
7	Blockage both sides	0.58					0.39
8	Blockage complete	0.49					0.34
9	Blockage bothered	0.58					0.46
10	Discharge a lot	0.79					0.57
11	Blow nose 10x daily	0.73					0.52
12	Discharge bothered	0.75					0.57
13	Cough lie down	0.49					0.34
14	Lump in throat	0.52					0.37
15	PND bothered	0.71					0.55
16	Smell loss complete				0.87		0.75
17	Smell loss bothered				0.71		0.52
18	Facial pain 5+		0.79				0.62
19	Facial pain bothered		0.83				0.68
20	Facial pressure severe		0.71				0.49
21	Facial pressure bothered		0.79				0.62
22	Headaches						0.12
23	Fever						0.09
24	Coughing			0.48			0.3
25	Bad breath						0.13
26	Fatigue						0.14
27	Nasal itching						0.15
28	Sneezing			0.31			0.22
29	Eye itching						0.22
30	Eye tearing						0.21
31	Ear fullness					0.64	0.43

32	Ear pain		0.63	0.41
33	Ear pressure		0.74	0.53
34	Wheezing	0.54		0.3
35	Chest tightness	0.59		0.34
36	Shortness of breath	0.57		0.33
37	Cold/flu symptoms	0.39		0.2

Table A4. Symptom commonalties from the exploratory factor analysis (EFA) of the 37 presence, severity, and secondary CRS symptom changes from baseline to 6 months and 6 months to 16 months. EFA was fit using ordinary least squares and an oblimin rotation (number of patients = 3535).

#	Item Label	Baseline – 6 month follow up	6 - 16 month follow up
1	Blockage	0.49	0.3
2	Discharge discolored	0.24	0.19
3	PND	0.45	0.28
4	Smell loss	0.42	0.47
5	Facial pain	0.51	0.47
6	Facial pressure	0.51	0.41
7	Blockage both sides	0.39	0.28
8	Blockage complete	0.34	0.23
9	Blockage bothered	0.46	0.4
10	Discharge a lot	0.57	0.5
11	Blow nose 10x daily	0.52	0.43
12	Discharge bothered	0.57	0.54
13	Cough lie down	0.34	0.29
14	Lump in throat	0.37	0.27
15	PND bothered	0.55	0.4
16	Smell loss complete	0.75	0.69
17	Smell loss bothered	0.52	0.48
18	Facial pain 5+	0.62	0.6
19	Facial pain bothered	0.68	0.63
20	Facial pressure severe	0.49	0.44
21	Facial pressure bothered	0.62	0.54
22	Headaches	0.12	0.14
23	Fever	0.09	0.1
24	Coughing	0.3	0.29
25	Bad breath	0.13	0.12
26	Fatigue	0.14	0.14
27	Nasal itching	0.15	0.19
28	Sneezing	0.22	0.25
29	Eye itching	0.22	0.29
30	Eye tearing	0.21	0.3
31	Ear fullness	0.43	0.41

32	Ear pain	0.41	0.33
33	Ear pressure	0.53	0.39
34	Wheezing	0.3	0.33
35	Chest tightness	0.34	0.4
36	Shortness of breath	0.33	0.37
37	Cold/flu symptoms	0.2	0.24

Chapter 4 - Conclusion

There are many uses for, and methods of, conducting EFA. In this thesis, I have proposed a new method to identify the number of factors to extract, studied its performance in application to certain data structures, and applied EFA model selection and factor extraction methods to estimate latent structure in symptoms common in CRS and its related co-morbid conditions.

The proposed method for determining the number of factors to extract during an EFA adds to the vast literature addressing the problem of estimating m and how to navigate this situation. This new m-estimation procedure performed well under a variety of simulated testing conditions which varied with regard to sample size (N), data dimensionality (P), and strength of correlation structure. Thus, this method may be a viable and versatile option of estimating the underlying factor model when sample size is sufficiently large.

The CRS symptom EFA shed light on the studied symptoms, which decomposed into five interpretable factors, generating several hypothesized biological factor underpinnings. We were able to identify congestion and discharge, smell loss, ear and eye, asthma and constitutional, and facial pain and pressure symptom factors. These factors are consistent with understanding of biology and pathological processes in individual sinuses.

Our CRS study utilized Cattell's scree test (5, 5, and 5 factors) and parallel analysis (5, 5, and 6 factors) in order to determine the number of factors to extract for the baseline, 6-month follow-up, and 16-month follow-up questionnaires. Interestingly, these methods estimated modestly different m compared with the Kaiser eigenvalue greater than 1 rule (K1; 6, 5, and 7 factors) and quite different *m* compared with other commonly utilized methods including the Bayesian information criterion (BIC; 15, 16, and 14 factors) and sample size adjusted BIC (SSBIC; 17, 17, and 20 factors), for baseline, 6-month, and 16-month follow-up questionnaires, respectively. Our newly proposed trace method also produced an optimal factor cardinality far removed from those presented in the CRS paper (13, 13, and 16 for baseline, 6-month, and 16-month questionnaires, respectively). In Chapter 2 we hypothesized that these differences may be explained by differing standards of fit implicated by the different levels of specificity

(dimensionality versus distributional form) addressed by the methods' objective criteria. Further research is needed to elucidate this conjecture.

Determining which method to utilize for m-estimation is difficult for several reasons. Firstly, there are a large number of potential options for estimating the number of factors with potentially different theoretical foundations including likelihood-based methods, eigenvalue-based methods, graphical methods, and cross-validated or bootstrap methods. Investigators must first consider the purpose of their analysis when deciding which method to utilize. If interpretability or conciseness is of paramount importance, one may consider methods aligned with this ideal. Otherwise, for example, if one is placing emphasis on identifying the number of factors in a FA model hypothesized to literally underlie the data, methods attuned to that goal such as BIC or TRACE should be considered. This target determination is important, as it will drive the results and inference downstream in the analysis. This thesis has shown that the choice of m can substantively impact qualitative and quantitative changes in loading and factor interpretations. As such, this choice directly influences whether the researcher's desired goal is attained with respect to unbiased estimation, verisimilitude, generalizability, or interpretability. The results are thus of high importance to researchers conducting EFAs.

We recommend that future work focus on the decision of which method(s) to use when attempting to find m in EFA settings. The best process of choosing which method of estimating m may very well be, firstly identifying what interpretation of m is relevant for the current study, narrowing the field of potential methods. Following this, a practitioner will likely still be faced with choosing between several methods which may perform differently in application to the observed data. It is clear from the simulation study that under certain, possibly identifiable conditions, methods may outperform or underperform compared to their average efficacy across conditions. Because the strength of correlations and sample size of observed data were strong drivers of the efficacy of comparative methods, these attributes along with others should shed light on which method is most appropriate. Thus, it might be that observed correlation matrix attributes could be utilized within a single analysis to determine which methods would perform best and future work in this area also would be valuable. A practitioner could then choose between methods with an understanding and anticipation of which methods

may be most appropriate for their specific data at hand. Finally, agreement between methods may prove to be evidence that the agreed upon m is desirable compared to other possibilities. Simulation studies such as the one described in the methods portion of the thesis can address these questions for us, by testing hypothesized methods against a known truth we generate.

This thesis was able to identify similar latent structure and factor identity in three CRS symptom questionnaire administrations, as well as the changes in symptom response scores between administrations. These EFAs were consistent with the hypothesis that hypothesized biopathological phenomena underlay the observed symptom responses. However, objective sinus inflammation data must be incorporated in order to adequately assess this hypothesis.

The trace method showed promise as a viable additional method for EFA model selection, outperforming many commonly utilized methods across several simulation conditions. However, in the diverse range of fields where EFA is utilized, the simulated scenarios were small in scope, as the number of factors assessed was always between 5 and 10, the number of variables utilized was between 11 and 100, and the number of simulated samples was between 100 and 1000. This thesis brings to light alternative approaches to EFA and EFA model selection that we hope will prove useful as they are further refined.

References

Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika*, 255–265.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, *52*(3), 317–332. https://doi.org/10.1007/BF02294359

Bai, J., & Ng, S. (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica*, *70*(1), 191–221. https://doi.org/10.1111/1468-0262.00273

Brown, T. A. (2014). *Confirmatory factor analysis for applied research*. Guilford Publications.

Browne, J. P., Hopkins, C., Slack, R., & Cano, S. J. (2007). The Sino-Nasal Outcome Test (SNOT): Can we make it more clinically meaningful? *Otolaryngology–head and Neck Surgery*, *136*(5), 736–741.

Buuren, S. van, & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, *45*(3), 1–67.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*(2), 245–276.

Chang, H., Lee, H. J., Mo, J.-H., Lee, C. H., & Kim, J.-W. (2009). Clinical implication of the olfactory cleft in patients with chronic rhinosinusitis and olfactory loss. *Archives of Otolaryngology–Head & Neck Surgery*, *135*(10), 988–992.

Cole, M., Schwartz, B., & Bandeen-Roche, K. (2017). Exploratory Factor Analysis of CRS Symptoms.

DeConde, A. S., Bodner, T. E., Mace, J. C., & Smith, T. L. (2014). Response Shift in Quality of Life After Endoscopic Sinus Surgery for Chronic Rhinosinusitis. *JAMA Otolaryngology–Head & Neck Surgery*, *140*(8), 712–719. https://doi.org/10.1001/jamaoto.2014.1045

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272–299. https://doi.org/10.1037/1082-989X.4.3.272

Ferguson, B. J., Narita, M., Yu, V. L., Wagener, M. M., & Gwaltney, J. M. (2012). Prospective Observational Study of Chronic Rhinosinusitis: Environmental Triggers and Antibiotic Implications. *Clinical Infectious Diseases*, *54*(1), 62–68. https://doi.org/10.1093/cid/cir747

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1). Springer series in statistics Springer, Berlin.

Hamilos, D. L. (2011). Chronic rhinosinusitis: Epidemiology and medical management. *Journal of Allergy and Clinical Immunology*, *128*(4), 693–707. https://doi.org/10.1016/j.jaci.2011.08.004

Hirose, K., Kawano, S., Konishi, S., & Ichikawa, M. (2011). Bayesian information criterion and selection of the number of factors in factor analysis models. *Journal of Data Science*, *9*(2), 243–259.

Hirsch, A. G., Stewart, W. F., Sundaresan, A. S., Young, A. J., Kennedy, T. L., Scott Greene, J., … Schwartz, B. S. (2017). Nasal and sinus symptoms and chronic rhinosinusitis in a population-based sample. *Allergy*, *72*(2), 274–281. https://doi.org/10.1111/all.13042

Hopkins, C., Browne, J. P., Slack, R., Lund, V., & Brown, P. (2007). The Lund-Mackay staging system for chronic rhinosinusitis: How is it used and what does it predict? *Otolaryngology - Head and Neck Surgery*, *137*(4), 555–561. https://doi.org/10.1016/j.otohns.2007.02.004

Hopkins, C., Browne, J. P., Slack, R., Lund, V., Topham, J., Reeves, B., … van der Meulen, J. (2006). The national comparative audit of surgery for nasal polyposis and chronic rhinosinusitis. *Clinical Otolaryngology: Official Journal of ENT-UK ; Official Journal of Netherlands Society for Oto-Rhino-Laryngology & Cervico-Facial Surgery*, *31*(5), 390–398. https://doi.org/10.1111/j.1749-4486.2006.01275.x

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. https://doi.org/10.1007/BF02289447

Humphreys, L. G., & Jr, R. G. M. (1975). An Investigation of the Parallel Analysis Criterion for Determining the Number of Common Factors. *Multivariate Behavioral Research*, *10*(2), 193–205. https://doi.org/10.1207/s15327906mbr1002_5

Kaiser, H. F. (1960). The Application of Electronic Computers to Factor Analysis. *Educational and Psychological Measurement*, *20*(1), 141–151. https://doi.org/10.1177/001316446002000116

Kamata, A., & Bauer, D. J. (2008). A note on the relation between factor analytic and item response theory models. *Structural Equation Modeling*, *15*(1), 136–153.

Lee, C.-T., Zhang, G., & Edwards, M. C. (2012). Ordinary Least Squares Estimation of Parameters in Exploratory Factor Analysis With Ordinal Data. *Multivariate Behavioral Research*, *47*(2), 314–339. https://doi.org/10.1080/00273171.2012.658340

Lopes, H. F., & West, M. (2004). Bayesian Model Assessment in Factor Analysis. *Statistica Sinica*, *14*(1), 41–67.

Myung, I. J. (2000). The Importance of Complexity in Model Selection. *Journal of Mathematical Psychology*, *44*(1), 190–204. https://doi.org/10.1006/jmps.1999.1283

Norris, M., & Lecavalier, L. (2010). Evaluating the Use of Exploratory Factor Analysis in Developmental Disability Psychological Research. *Journal of Autism and Developmental Disorders*, *40*(1), 8–20. https://doi.org/10.1007/s10803-009-0816-2

Owen, A. B., & Wang, J. (2016). Bi-Cross-Validation for Factor Analysis. *Statistical Science*, *31*(1), 119–139. https://doi.org/10.1214/15-STS539

Preacher, K. J., Zhang, G., Kim, C., & Mels, G. (2013). Choosing the optimal number of factors in exploratory factor analysis: A model selection perspective. *Multivariate Behavioral Research*, *48*(1), 28–56.

Press, S. J., & Shigemasu, K. (1999). A note on choosing the number of factors. *Communications in Statistics-Theory and Methods*, *28*(7), 1653–1670.

R Core Team. (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Revelle, W. (2017). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Evanston, Illinois: Northwestern University. Retrieved from https://CRAN.R-project.org/package=psych

Schwarz, G., & others. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464.

Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, *52*(3), 333–343.

Sieskiewicz, A., Lyson, T., Olszewska, E., Chlabicz, M., Buonamassa, S., & Rogowski, M. (2011). Isolated sphenoid sinus pathologies–the problem of delayed diagnosis. *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research*, *17*(3), CR179.

Sundaresan, A., Hirsch, A., Young, A., Tan, B., Schleimer, R., Kern, R., … Schwartz, B. (2017). Longitudinal Evaluation of Chronic Rhinosinusitis Symptoms in a Population-based Sample.

Swihart, B. J., Caffo, B., James, B. D., Strand, M., Schwartz, B. S., & Punjabi, N. M. (2010). Lasagna plots: a saucy alternative to spaghetti plots. *Epidemiology (Cambridge, Mass.)*, *21*(5), 621.

Tan, B. K., Kern, R. C., Schleimer, R. P., & Schwartz, B. S. (2013). Chronic Rhinosinusitis: The Unrecognized Epidemic. *American Journal of Respiratory and Critical Care Medicine*, *188*(11), 1275–1277. https://doi.org/10.1164/rccm.201308-1500ED

Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods*, *16*(2), 209–220. https://doi.org/10.1037/a0023353

Tustin, A. W., Hirsch, A. G., Rasmussen, S. G., Casey, J. A., Bandeen-Roche, K., & Schwartz, B. S. (2017). Associations between Unconventional Natural Gas Development and Nasal and Sinus, Migraine Headache, and Fatigue Symptoms in Pennsylvania. *Environmental Health Perspectives*, *125*(2), 189–197. https://doi.org/10.1289/EHP281

Underwood, L. G., & Teresi, J. A. (2002). The daily spiritual experience scale: development, theoretical description, reliability, exploratory factor analysis, and preliminary construct validity using health-related data. *Annals of Behavioral Medicine*, *24*(1), 22–33. https://doi.org/10.1207/S15324796ABM2401_04

Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct Explication through Factor or Component Analysis: A Review and Evaluation of Alternative Procedures for Determining the Number of Factors or Components. In R. D. Goffin & E. Helmes (Eds.), *Problems and Solutions in Human Assessment* (pp. 41–71). Springer US. https://doi.org/10.1007/978-1-4615-4397-8_3

Wald, E. R., Milmoe, G. J., Bowen, A., Ledesma-Medina, J., Salamon, N., & Bluestone, C. D. (1981). Acute maxillary sinusitis in children. *New England Journal of Medicine*, *304*(13), 749–754.

Wj, F., Vj, L., J, M., C, B., I, A., F, B., … Pj, W. (2012a). EPOS 2012: European position paper on rhinosinusitis and nasal polyps 2012. A summary for otorhinolaryngologists. *Rhinology*, *50*(1), 1–12. https://doi.org/10.4193/Rhino50E2

Wj, F., Vj, L., J, M., C, B., I, A., F, B., … Pj, W. (2012b). European Position Paper on Rhinosinusitis and Nasal Polyps 2012. *Rhinology. Supplement*, (23), 3 p preceding table of contents, 1-298.

Zwick, W., & Vejicer, W. (1984). A Comparison of Five Methods for Determining the Number of Components in Data Sets. *Psychological Bulletin*, *99*(3).

Biography

Matthew K. Cole was born in 1993 in the USA.

Matt completed his undergraduate work at Sacred Heart University in Fairfield, Connecticut, where he majored in Biology and Mathematics. During his undergraduate education, he spent some time studying abroad in Italy and Germany and spent his other summers researching the ecology of the American Horseshoe Crab *Limulus polyphemus* in Long Island Sound.

In 2015, Matt began his Sc.M. at Johns Hopkins University. He was a teaching assistant for the Statistical Methods in Public Health course sequence.