Exploring Persistent Throat Symptoms

James Thomas O'Hara

MD

Population Health Sciences Institute

Newcastle University

Submitted September 2020

Abstract

Background

Patients commonly present with a range of persistent throat symptoms. There is still much research to be done to understand how the individual symptoms relate to each other. An improved understanding of any symptom relationships could aid in identifying groups of patients for particular treatments. An opportunity to explore throat symptoms in detail was provided by a clinical trial to assess the effectiveness of stomach acid lowering medication on patients' throat and voice complaints. The aim was to identify a clinically meaningful classification of patients' symptoms.

Methods

Baseline data for all patients entering the Trial of Proton Pump Inhibitors in Throat Symptoms (TOPPITS) was provided by the Newcastle University Clinical Trials Unit. Data included: demographics, three separate symptom questionnaires and a scoring assessment of throat appearances. The relationships between patient demographics, symptom scores and throat appearances were explored with scatter plots and pairwise correlation. Exploratory factor analysis (EFA) of the combined symptom questionnaires was conducted. Cluster analysis of patients using the factors generated by the EFA was then performed.

Results

Data for 344 patients analysed. The distributions of the questionnaire scores were comparable with published literature. The total scores from the three questionnaires were positively related with each other. No relationship was observed between the throat appearances and any of the three symptom questionnaires in this population. The EFA led to a seven factor model comprising factors of: voice, cough, gastrointestinal symptoms, airway symptoms and dysphagia, throat clearing, life events, and lump in the throat sensation. Cluster analysis failed to identify clinically meaningful groups of patients.

Conclusions

The TOPPITS baseline data confirmed that patients recruited to the trial reflect the wider population of patients presenting with persistent throat symptoms. No evidence of an association between throat appearances and patient reported symptoms was found in this

study. Dimension reduction offered a simplified classification of symptoms, but clusters of patients based on this classification could not be identified. These results imply that individual throat symptoms cannot be used to define patient groups and that the term “persistent throat symptoms” to encompass all symptoms is appropriate to use in clinical practice.

Acknowledgements

I would like to thank, in particular, Professor Janet Wilson for her encouragement in pursuing this thesis. Professor Wilson has been my research mentor for over 14 years and has always offered me superb advice and guidance. I am extremely grateful to her.

I would also like to thank my thesis supervisors, alongside Professor Wilson, Dr Louise Hayes and Dr Holly Fisher. They have provided excellent advice and support, in particular regarding the statistical approaches utilised within this thesis.

This thesis arose from the TOPPITS randomised controlled trial. The TOPPITS Trial Management Group were helpful in defining how this thesis work aligned with a governed trial of an interventional medicinal product. The data used in this thesis has been collected within 8 UK hospitals. I would like to thank the patients, who volunteered to enter the clinical trial, and the research teams in each centre for contributing to the trial.

Finally, I would like to thank my wife and two sons, who have continued to encourage me in conducting this thesis.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Definition
AIC	Akaike Information Criteria
ANCOVA	Analysis of Covariance
CI	Confidence Interval
CReSS	Comprehensive Reflux Symptom Score
DMC	Data Monitoring Committee
ENT	Ear Nose and Throat
EOR	Extra Oesophageal Reflux
GORD	Gastro-oesophageal Reflux Disease
GP	General Practitioner
HTA	Health Technology Assessment
IMP	Investigational Medicinal Product
IQR	Interquartile Range
LPR	Laryngopharyngeal Reflux
LPR HRQL	Laryngopharyngeal Reflux – Health Related Quality of Life
NCTU	Newcastle Clinical Trials Unit
PPI	Proton Pump Inhibitor
p	Probability
RFS	Reflux Finding Score
RSI	Reflux Symptom Index
RSI-HB	Reflux Symptom Index minus the heartburn/dyspepsia item
sd	Standard deviation
SBC	Schwartz's Bayesian Criterion
SE	Standard Error
TMG	Trial Management Group
TOPPITS	Trial Of Proton Pump Inhibitors in Throat Symptoms
TSC	Trial Steering Committee
UK	United Kingdom
USA	United States of America

Chapter 1 Introduction

1.1 Background

"Persistent throat symptoms" comprise a group of interlinked complaints that include a hoarse voice, a feeling of a lump in the throat – often termed "globus" sensation, the need to clear one's throat, mucus in the throat or catarrh, cough and throat discomfort. Patients with these symptoms are often referred to secondary care for assessment and management. In 2010, 2.7million patients in England were referred to Ear, Nose and Throat (ENT) outpatient clinics for assessment of their medical symptoms[1]. In a study of over a thousand middle aged women, six percent reported a feeling of a lump in the throat[2]. In a similar sized, throat - specific questionnaire study of all GP attenders to a single practice, 27% of patients reported abnormally high throat symptoms[3]. The author of this thesis, and colleagues, have suggested a conservative estimate of 60,000 patients referred each year to secondary care with persistent throat symptoms in England[4]. However, based on the data presented above and assuming five percent of the 2.7 million referrals have persistent throat symptoms, the estimate would be more than double, and closer to 135,000 referrals in England per year. Thus, management of these symptoms presents a significant burden to health care providers.

Persistent throat symptoms have proven difficult to manage for both medical practitioners and patients. Often there is no obvious underlying cause for the symptoms that can help target therapy. As such, these complaints would reasonably fall into the category of "persistent physical symptoms", which may have been included previously in terms such as "medically unexplained symptoms". Patients with persistent throat symptoms are referred to ENT for a number of reasons; the main one being that specialist endoscopy of the larynx and pharynx is available. Most general practitioners (GPs) assessing these symptoms will recognise the chronic nature of the symptoms and often the intermittent occurrence of symptoms, and conclude that an underlying physical abnormality is unlikely. However, to rule out a treatable cause for the symptoms, endoscopy is required, and is usually only available in ENT or some speech and language therapy (SaLT) clinics. Treatable causes for throat symptoms can usually be predicted by the nature of the symptoms. For instance, voice quality which is always hoarse and does not fluctuate in nature is far more likely to be due to a physical, abnormal lesion on the vocal cord, than voice quality that varies in nature, for instance after talking for a period of time. The latter is more likely to be "functional" in nature and theories postulate that tension or spasm in the muscles of the voice box may contribute to symptoms.

There is a known association between functional voice symptoms and psychological symptoms such as anxiety and perfectionism, and also with chronic fatigue[5]. These patients benefit from speech therapy delivered through specialist SaLT services[6].

In recent years the referral pattern of patients with persistent throat symptoms has changed. There has been increasing emphasis from the Department of Health (DoH) towards early referral for patients with symptoms that could suggest underlying malignancy. Given the vague nature of persistent throat symptoms, the difficulty in examining the throat well for GPs and the lack of a definable underlying cause for these symptoms, many patients are now referred on a two-week rule pathway; patients must be seen in specialist clinics within twoweeks of the GP referral. The defined terms that should prompt such a referral are those of persistent hoarse voice, difficulty swallowing and pain in the throat. It is understandably difficult for GPs, who lack the ability to examine the throat, to separate chronic, perhaps "functional" type symptoms, from those that may be more concerning for an underlying throat cancer. The DoH rightly wants more patients with cancer identified early to improve survival outcomes. One method to do this is to assess more patients urgently. The two-week-wait referral pathway has been recommended not just for patients who are felt to have cancer, but for patients whose symptoms may suggest cancer. The aim is to identify three percent of referrals who have an underlying cancer; symptoms that merit referral should have a minimum three percent positive predictive value in detecting cancer[7]. To do this requires seeing 97% of patients who do not. It is for this reason, that patients with persistent throat symptoms are now often referred to exclude cancer.

1.1.1 *The symptoms*

There are a number of symptoms grouped together to form "persistent throat symptoms", as defined for this thesis. There is no currently accepted terminology as such for these symptoms. Whilst the symptoms may sound distinct in nature, the reality is that there is significant overlap between them and highlighting one symptom over another by either the patient or clinician will have a large element of subjectivity associated to the labelling of the complaint. The most well-known "medically unexplained symptom" in the throat is that of "a feeling of a lump", and has been noted to be reported by up to 45% of otherwise healthy people[8]. It is claimed[9] that the term *globus hystericus* was first described in 1707 by Purcell[10]; originating from the Latin *globus* meaning globe and *hystericus* presumably to indicate a psychological component to the symptom. It was later adopted into the Oxford English dictionary. Since then the symptom has been referred to as *globus* alone or *globus pharyngeus[11].* Clinicians have sought physical explanations for the symptoms for many

years, suggesting amongst other causes: tightness of the upper oesophageal sphincter (cricopharyngeus), bony growths of the cervical spine, iron-deficiency anaemia, and more recently gastroesophageal reflux (GORD)[9].

Recently, Doody and Fenton have questioned whether a more contemporary term should be adopted to describe globus pharyngeus[12]. They have suggested “troublesome throat awareness”, stating that a feeling of a lump in the throat on dry swallowing is a sensation that everyone can become aware of. Our research team has used the term “chronic throat symptoms”. However, having discussed this with an academic psychologist, we were recommended to use the term persistent throat symptoms, as a label more acceptable to patients.

Whilst globus sensation has received the most interest, this may reflect that it is the medically unexplained symptom most connected to the throat. Rhinologists are well aware of the difficulty in managing catarrh or “post-nasal drip”[13], and pulmonologists have sought to define algorithms for managing chronic cough[14]. In the absence of identifiable underlying chronic sinusitis or chronic airways disease, these two symptoms may well overlap with globus. The repeated need to clear the throat is often associated with globus, but equally may occur with the sensation of mucus coming into the throat from the nose, and one person’s throat clearing is another person’s cough. Catarrh has been identified to be a general throat symptom, rather than being associated with either globus or difficulty swallowing[2].

The link between persistent throat symptoms and psychological distress is well documented. Patients who complain of globus sensation report higher levels of anxiety and depression. The degree of distress caused by throat symptoms was found to be independently predicted by patients’ anxiety levels, in a study of 105 patients[2]. Patients with globus sensation reported higher rates of significant life events in the preceding year, than controls (for example relationship / marital breakdown). They also reported fewer confiding relationships with partners than controls[15]. Female patients with dysphonia reported higher levels of psychological distress and more medically unexplained symptoms than healthy controls and GP attenders[16]. In a study that included over 4000 male US veterans, 6.4% reported globus sensation. These patients tended to be from a lower socioeconomic background, and had an increased risk of being diagnosed with major depression, general anxiety, post-traumatic stress disorder and other persistent physical symptoms – when compared to patients who did not report globus[17].

1.1.2 *Gastroesophageal reflux as a cause for chronic throat symptoms.*

The prevalence of gastroesophageal reflux disease (GORD) has been estimated at 10-20% of the Western World population, as defined by at least weekly heartburn or acid regurgitation, based on a systemic review of epidemiology studies in 2005[18]. The concept that backflow of stomach acid leads to throat symptoms, has been debated for over 30 years. In addition to the proposal that gastric acid damages the mucosa of the throat and voice, leading to symptoms, other causes have been set out such as the theory that gastric enzymes and bile acid, as components of the gastric refluxate, could directly damage the throat lining. In vivo experiments have demonstrated tissue level damage[19]. Cellular changes have been noted in response to individual and grouped components of gastric juice (acid, pepsin and bile) on laryngopharyngeal cells, with gene expression changes, immunohistochemical effects and changes in molecular regulator activity demonstrated[20]. Some of these cellular changes have also been demonstrated in laryngeal samples taken from patients with persistent throat symptoms[20].

The study most frequently cited to have popularised the theory that reflux of stomach contents leads to throat and voice symptoms was described by Koufman in 1991[21]. This paper introduced the term “silent reflux”; two hundred and twenty five patients with throat and voice symptoms underwent pH monitoring of the distal and upper oesophagus. Less than half complained of heartburn or acid regurgitation; hence over half were deemed to suffer with silent reflux. Sixty-two percent of all patients had abnormal oesophageal pH recordings. Hoarseness, globus and chronic cough were frequent symptoms in this patient population. Since then, Koufman and colleagues have publicised the theory, developed patient reported outcome measures [22] and a throat appearance scoring tool[23], and advocated the use of acid reduction medications in the form of proton-pump inhibitors (PPIs). ENT clinicians readily took up this theory as a treatment for hitherto resistant symptoms, with a survey in 2007 of UK based otolaryngologists showing 90% believed in the condition and over 50% prescribed PPIs[24]. Multiple case series showed patients’ throat and voice symptoms responded over time with this treatment. However, randomised clinical trials did not support the now widespread clinical practice, failing to demonstrate any benefit of PPIs over placebo[25].

In many ways analogous to chronic cough algorithms that advocated empirical trials of antihistamines, so patients with throat and voice symptoms started to be treated with empirical trials of PPIs; a practice often described in GP referral letters, with failure to respond seemingly justifying a specialist referral. Along with this practice came a term for the

condition – “Laryngopharyngeal Reflux”, or LPR. Other terms such as extraoesophageal reflux or supraoesophageal reflux have been used in the literature.

The diagnosis of reflux as a cause for throat symptoms relies upon symptoms, signs, investigations and treatment outcomes. Many clinicians sceptical of LPR as an entity will cite the lack of coherence between the factors required to make the diagnosis. Symptoms and throat signs may correlate at presentation[26], although very little evidence exists in this area and the study cited included only 40 patients with a high rate of throat signs seen. Observational studies of large series of patients show that following treatment with PPIs, patients’ symptoms and throat appearances improved[27, 28]. Objective diagnostic tests, in particular oesophageal pH monitoring, have been used in a number of smaller observational studies to “diagnose” LPR by identifying abnormal acid exposure in the upper oesophagus over 24 hours. Some evidence does exists to demonstrate that acid can enter the throat; more reflux events and greater pharyngeal acid exposure times were recorded in the laryngopharynx of people with throat symptoms compared to asymptomatic controls in a large meta-analysis (n=793), using 24-ambulatory pH monitoring[29]. However, pH monitoring results do not appear to correlate with patients’ symptoms [30]. Finally, high quality evidence is lacking in support of PPIs to treat symptoms over and above offering a placebo effect. This does not mean that reflux does not play a role in throat symptoms, but reflects the reliance on PPIs in treating these patients and the vast bulk of research performed in this field relating to these medications. Very little evidence exists assessing the role of other factors which may reduce reflux, such as diet[31], alginates[32] and lifestyle[33].

A recent literature search revealed six systematic reviews of randomised controlled trials, assessing the evidence for PPIs in throat symptoms [25, 34-38]. Most trials included were of less than 100 patients, with the largest study containing 145 patients. Overall the reviews do not demonstrate any improvement of symptoms with treatment over placebo, however two of the reviews specifically assessed the response in trials that reported the Reflux Symptom Index (RSI) and concluded that an improvement in pooled RSI scores was observed[34, 36]. It is clear from all of these studies that a significant placebo effect exists for this condition, as symptoms improved similarly in both the PPI and placebo groups.

Reviewing the systematic reviews and meta-analyses of the trials assessing response to PPIs has caused some concern over the methodological rigour employed. Wei[36], as a sole author, concluded that PPI treatment could improve symptoms in patients with LPR. Within the text in figure 3, comparing the Reflux Symptom Index outcomes, is included a study by Shaheen

et al[39]. On closer examination of the original manuscript, this controlled trial did not collect the Reflux Symptom Index as an outcome. A recent electronic publication by Lechien et al [40]concluded a mild superiority of PPIs over placebo. Within their meta-analysis of placebo controlled trials is a trial by Ezzat et al[41], which compared a prokinetic agent and PPI versus PPI alone. We have contacted the authors to suggest this trial should not have been included in the stated meta-analysis as it assesses the role of the prokinetic agent and not the PPI. There remains a great interest in "LPR", as evidenced by the popularity in prescribing patients PPIs for throat symptoms, and the ongoing publication of further meta-analyses.

1.2 The Trial of Proton Pump Inhibitors in Throat Symptoms

Section 1.1 has described the clinical condition and background to patients with persistent throat symptoms. Section 1.2 will explain the basis behind the design of a large clinical trial, exploring the treatment of these symptoms. This clinical trial has provided the data utilised in the analysis for this thesis. Section 1.4 will describe the patient reported outcome measures used and the assessment used to examine the throat of patients.

In 2011 the National Otolaryngology Trials Office, led by Professor Wilson, organised a conference entitled "Extraoesophageal Reflux – throat symptoms and cough: separating myth and evidence"[42]. I was invited to present an evidence synthesis at this meeting, summarising the use of alginates (eg. Gaviscon) and older formulations of stomach acid reduction medications. Following the conclusions from this conference, a clinical research theme commenced. It was clear that the evidence was lacking to support the current practice of acid-reduction prescriptions for throat and voice symptoms. The organising team concluded that "studies assessing proton-pump inhibitors (PPIs) in extra-oesophageal reflux suffer from variable study design and quality, small numbers and heavy selection bias and use a variety of different treatment regimens". A clinical trial was conceived with the Newcastle University Research Design Service in 2012, initially aiming to assess the clinical effectiveness of a PPI and alginate. Following a review by the Health Technology Assessment of the National Institute for Health Research, funding was achieved in 2013 to conduct a randomised double-blind placebo controlled study to assess the clinical effectiveness of a PPI (lansoprazole). The inclusion of an alginate in the study was removed following the funding review. The trial was entitled the Trial of Proton Pump Inhibitors in Throat Symptoms, or TOPPITS (trial registration numbers: ISRCTN38578686, EudraCT number 2013-004249-17). I was involved in the study design from the outset as a co-applicant and have been a Trial Management Group (TMG) member since the trial commenced. I have acted as Principal Investigator on the trial when I worked in Sunderland, and have recruited patients to the trial

since 2015 in Newcastle. In addition to contributing to the trial delivery and TMG oversight, I have taken an active role in the subsequent analysis of results. The following description of the TOPPITS trial is a summary taken from the published trial protocol [43].

1.2.1 *TOPPITS Protocol*

Aim

This was a non-commercial study to determine the clinical effectiveness of the proton pump inhibitor lansoprazole compared with placebo, in patients referred to secondary care with persistent throat symptoms.

Inclusion Criteria

Referred with a persistent (over 6 weeks) primary throat symptom—globus, hoarseness, throat clearing, throat discomfort, choking spasms, excess mucus/ postnasal drip or otherwise unexplained night-time cough or choking. Score of 10 or more on the non-heartburn items of the RSI.

Outcomes Measures

Three patient-reported outcome measures were collected, along with a score of the voice box appearances, and basic demographic details. The measures were taken at baseline (entry to the trial), four months after treatment and 12 months.

The TOPPITS primary outcome was the change in Reflux Symptom Index (RSI) at four months in the treatment and placebo groups in an intention to treat analysis. The nine-item RSI total score (0-45) allowed comparison with previous studies. We planned however to report also the RSI score omitting the heartburn item (0-40), which we and others noted can skew the results in favour of PPI in past small trials. Our proposed analysis would also address the issue of severity variation, through our stratification variables of site and baseline severity.

The Comprehensive Reflux Symptom Score (CReSS) is a 34 item questionnaire of oesophageal and extra-oesophageal symptoms[44]. It has been tested on groups of "throat" patients, healthy controls and those attending for an upper gastrointestinal endoscopy. It has three statistically robust symptom factors: gastrointestinal/oesophageal (14 items), an upper airway factor (eight items) – relating to cough, breathing, mucus and hoarseness, and a third, obstruction / choking globus factor (pharyngeal factor- seven items)[45].

Quality of life: Change in Laryngopharyngeal Reflux-Health Related Quality of Life (LPR-HRQL) total score and subscale at four and 12 months.

Laryngeal mucosal changes – as scored with the Reflux Finding Score (RFS).

The outcome measures which were used in the TOPPITS trial are shown in the Appendices. Subsequent chapters will detail the relevant published data and analyses of previous trials to have used the specific outcome.

Intervention

The active intervention was a 16-week (four-month) course of a 30mg twice daily dose of the proton pump inhibitor lansoprazole.

Control

Participants in the control arm received a 16-week (four-month) course of twice daily matched placebo capsule.

1.3 Methods to assess outcomes measures used for throat symptoms

Section 1.2 has described the TOPPITS trial, from which this thesis data arises, and has set out the patient reported outcome measures used and the scoring tool for the throat appearances. This section will explore some of the previous research into these outcome measures and the statistical analyses used. The chapter will also explore relevant statistical analyses that are in use to assess the utility of patient reported outcome measures, in patients with persistent throat symptoms and in other similar diseases.

1.3.1 *Patient reported outcome measures used in "LPR"*

Given the lack of any reproducible investigations to diagnose reflux related throat symptoms, or "LPR", and the lack of correlation between throat appearances and reflux investigations, patient reported outcome measures (PROMs) have become the principal method of assessing and monitoring throat symptoms[46]. A literature review using Ovid Medline was performed using the search terms "patient reported outcome measures" and "laryngopharyngeal reflux". This review intended to highlight the evidence pertaining to the three PROMs and the RFS used in TOPPITS, and to review the statistical methods used to analyse the PROMs in this setting.

Reflux Symptom Index

Originally validated in 2002, the RSI was assessed in 25 patients with LPR compared to 25 healthy controls[22]. The mean age of the LPR group was 57. The RSI was completed twice prior to LPR treatment to assess test-retest reliability. Results were compared to the 30-item Voice Handicap Index. LPR was "diagnosed" using dual-probe 24-hour pH monitoring. Patients were treated with 6 months twice daily PPI. Patients were not involved in the construction of the questionnaire. The mean RSI in the LPR group was 19.9 (sd. 11.1) and 20.9 (9.6) in the two pre-treatment tests. There was a mean 8(4) days between visits to complete the RSI. Test-retest reliability was r=0.81 using Pearson's correlation coefficient. After 6 months treatment, the mean RSI was 12.8 (10.0). The mean RSI in the control group was 11.6 (95% CI 9.7 – 13.6).

CRESS

The 34-item questionnaire was originally described in 2009[44] to include the RSI items and the Gastro-oesophageal reflux disease (GORD) Symptom Assessment Scale (GSAS). The GSAS was identified as an adequate symptom scale when assessed in a systematic review, although this review highlighted some deficiencies in all scales and the need for a more appropriate tool to use to measure therapies outcomes[47]. There was recognition that the RSI, whilst popular, lacked a throat pain symptom, and the traditional GORD symptoms were bundled in one poly-symptomatic item. The RSI item covering "heartburn, indigestion, chest pain and stomach acid coming up" was unbundled into four individual items. The 15 items from the GSAS were added, along with the 10 symptoms that had been originally included with the GSAS. These 10 items had been removed largely due to low endorsement via a telephone interview; the CReSS authors felt some of these items had been removed despite high endorsement and chose to include them all. Items common to both questionnaires were included as one. Each item is scored from zero to five. The questionnaire was given to 62 patients attending the ENT department with chronic throat symptoms, based on the referral letter.

Item endorsement (scores >zero) showed the following symptoms were recorded by over 50% of the patients: Throat clearing (RSI item, 79%), pressure / lump in throat (GSAS item, 70%), excess mucus (RSI, 66%), feeling things stuck in throat (GSAS item, 62%), back pain (GSAS omitted item - 59%), hoarseness (GSAS item, but is also on RSI, 52%), flatulence (GSAS item, 52%), mucus dripping down back of throat (RSI in addition to excess mucus, 51%), pain in throat (GSAS item, 51%).

Correlation of the disaggregated RSI items, within the CReSS questionnaire, was performed using Pearson's correlation coefficient[44]. Each RSI item contained significantly correlated disaggregated symptoms, other than pressure in chest which was not correlated with either indigestion or heartburn. Heartburn and indigestion were endorsed by 46%, and pressure in chest by 26%. Stomach acid coming up was reported by 36%. A cluster analysis was performed to assess if defined clusters of symptoms could map to clinical conditions. However, no new symptom patterns were identified and the methodology was not explored further. Whilst the specific type of cluster analysis used was not specified, the results state a dendrogram was evaluated within SPSS. This would suggest that the hierarchical cluster analysis methods had been used.

The 34-item questionnaire was subsequently termed the Comprehensive Reflux Symptom Scale (CReSS). It was further evaluated on 639 patients: 103 controls (staff, friends and hearing aid clinic attenders), 359 patients attending for oesophagogastroduodenoscopy (OGD), and 177 Ent patients referred with throat symptoms[45]. There was a high level of positive scoring (>zero) for all items on the CReSS. Interestingly, controls most frequently endorsed back pain (43%), headache (31%) and flatulence (26%); OGD patients heartburn (76%), indigestion (75%), and stomach acid coming up (72%) – i.e. three of the four symptoms in the single RSI item; and the ENT / throat patients throat clearing (80%), hoarseness (71%) and pressure / lump (65%). Internal reliability was assessed with Cronbach's α (0.93). A factor analysis using principal components analysis identified statistically robust symptom factors: gastrointestinal (oesophageal); an upper airway factor – relating to cough, breathing, mucus and hoarseness; and a third, obstruction / choking / globus factor (pharyngeal or throat)[45]. Each item is scored on a zero to five Likert scale, giving a range of total scores 0-170. There are 17 oesophageal items (score 0 – 85), nine upper airway items (score 0 – 45), and five pharyngeal items (score 0 – 25). Three items do not lie within the three main groups (score 0-15).

LPR-HRQL

The Laryngopharyngeal Reflux Health Related Quality of Life Questionnaire (LPR-HRQL) is a 43 item self-administered questionnaire[48]. It is composed of four domains (Voice – 12 items, Cough – six items, Throat Clearing – six items, and Swallow – five items). Each of these items is scored on a zero to six Likert scale. The scoring system for the questionnaire is described in section 2.2. The reliability, validity and responsiveness of this instrument was

assessed in 2005[48]. To date this questionnaire has not been widely used in studies. The rationale for its development was the recognition that health related quality of life instruments existed for GORD and voice, but not for "LPR". The instrument was developed from a literature review with input from several specialist clinicians.

The instrument was assessed on 117 patients diagnosed with LPR based on symptoms and throat / voice box appearances (RFS), who received twice daily PPIs for six months. Symptoms were assessed at two monthly time points including baseline. Patients also completed the Short-form 36 (SF-36), the Voice Handicap Index (VHI), the Quality of Life in Reflux and Dyspepsia (QOLRAD) and the Overall Treatment Effect (OTE). Statistical, construct and concurrent validity were analysed through descriptive statistics, factor analysis and relationships with the other questionnaires. Cronbach's α was used to assess internal reliability, and Pearson's correlation coefficient to assess test-retest reliability. Responsiveness was assessed over two time points with a paired t-test. The burden of completing the questionnaire was assessed by the ability and time to complete it.

The results stated good variability on all items with ranges spanning all potential values for each item. There were no floor or ceiling effects noted. The scales demonstrated normal distributions. All but one domain showed a single factor structure with the voice/ hoarse domain having more than one. The variability compared favourably with the VHI and QOLRAD. The voice/ hoarse domain correlated substantially (0.88) with the VHI (as would be expected, with both tools assessing similar symptoms). Internal consistency ranged from 0.84 – 0.93, although the swallow domain was lower at 0.69. The test-retest reliability ranged from 0.90 to 0.64. The responsiveness analysis suggested a minimum clinically meaningful difference of five on each overall domain score, and 10 for the overall impact of acid score. The questionnaire was felt to have been completed within 30 minutes.

Instrument validation in throat symptoms

The TOPPITS trial was conceived in 2012 and funded in 2013. The research team recognised certain deficiencies in the RSI, as the "market-leading" patient reported outcome measure and wished to include further symptom scales in the trial. Since the trial began further validation work had been conducted on the CReSS scale, as outlined above. A recent systematic review of PROMs relating to "LPR" performed in-depth analysis of each measure's development and

validity[46]. They identified seven relevant PROMs, including the RSI, CReSS (termed the LPR-34 in this review) and LPR-HRQL. They assessed each PROM against a checklist of six measurement properties[46] that comprised: conceptual model, content validity, reliability, construct validity, interpretability and scoring, burden and presentation. Unfortunately, despite the rigorously conducted review using transparent search criteria, the authors did not identify the further work on the CReSS questionnaire in July 2015, despite it having already been published in February 2015. As such, the LPR-34 as it was labelled, failed to meet many of the desired measurement properties. Had the more recent work been available to the authors then the CReSS questionnaire would have compared favourably in terms of reliability and construct validity.

The review does not single out a particular questionnaire as being the most rigorously developed. It recommends that over 100 participants be required for optimal factor analysis; only the LPR-HRQL of the three measures used in TOPPITS met this criterion on their available evidence (the subsequent CReSS analysis on 639 patients would have met this criterion). Chapter 4 will explain that the numbers of patients required for factor analysis also depends on the number of questionnaire items included; the more items, the more patients required. The RSI was developed on only 25 patients with the clinical condition of interest. The population on which the measures were assessed varied. Of course, defining such a population, when a definition of the condition does not exist is contentious. There are no established investigations that define LPR, hence the rationale to base treatments around patients' symptoms. With this in mind, it was clear that patient involvement in the development of measures was very much lacking. The review states that the LPR-HRQL engaged patients in its development, but the methods state a focus-group of clinicians were involved. Only the RSI and LPR-HRQL measured aspects of response to change, albeit in uncontrolled cohorts of patients treated with PPIs. The RSI was not assessed with a factor analysis as part of the validating work, however Printza et al[49] and Cathcart et al [50] did so in later studies. This work will be explored further in Chapter 4.

Exploratory factor analysis techniques have been used to analyse PROMs in throat symptoms as detailed above. Further details will be specified in Chapter 4 and Chapter 5. Other than the original study which set out the 34 –item questionnaire which led to the CReSS, cluster analysis has not been explored[44]. Factor analysis is a recognised item reduction technique which can then be used to define the variables included in cluster analysis. Adnane et al [51] assessed 131 patients undergoing treatment for chronic rhinosinusitis, collecting demographic data, objective investigations and a patient reported outcome measure. Three distinct clusters

of patients were identified with differing quality of life outcomes after treatment. The variables that were included in the cluster modelling were chosen following factor analysis that reduced the demographic data, comorbidity, subjective and objective clinical findings down to nine factors. A two-step cluster analysis was then performed using SPSS.

Reviewing the literature pertaining to the statistical analysis of throat symptoms PROMs and similar medical conditions has shown that exploratory factor analysis of multiple measures, and cluster analysis using the factors generated from factor analysis are recognised techniques. These techniques could reduce the large volume of symptom data obtained from TOPPITS into a more clinically meaningful group of symptoms.

1.4 The research question

At the outset of this thesis I chose to explore in detail the wealth of symptom information that the TOPPITS trial would offer. I set out to explore symptoms independent to treatment outcomes in the trial. TOPPITS was designed to address the common practice of PPI prescriptions for patients with throat symptoms. In the course of writing this thesis the TOPPITS results have been analysed and show no evidence of benefit in treating patients with PPI over placebo when PROMs were analysed. No trends towards a treatment effect have emerged. The one conclusion that readers will take from the trial is that PPIs should not be prescribed routinely for these patients. In truth, the trial's research team did anticipate this result given the lack of benefit demonstrated in smaller clinical trials. However, we did not perhaps envisage such a conclusive lack of treatment effect.

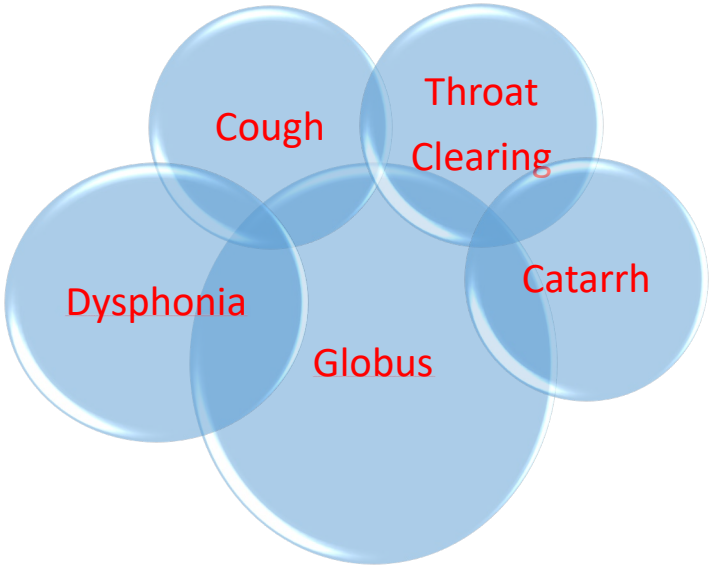
The work generated from this thesis has therefore gained greater importance. If acid suppression medications are not to be used for these patients, then how should they be treated? A greater understanding of the symptoms that patients present with would be very useful to explore other aetiologies and management strategies for these patients. The TMG has discussed whether subgroup analysis of the TOPPITS data would be legitimate. Defining subgroups based on an improved understanding of symptoms would be a pre-requisite to even consider such a secondary analysis.

Specific throat symptoms have often been investigated as single entities with research addressing globus, catarrh, post-nasal drip, dysphonia and cough alone. The concept of LPR has brought all these terms together, often through the patient reported outcome measures used in this condition. I believe there is significant overlap amongst the individual symptoms. Figure 1-1 shows a diagram I devised for a medically unexplained symptoms conference in 2013. The diagram's message was purely based on experience and anecdote, at the time.

TOPPITS gives the opportunity to look in detail at symptom reporting by patients and to assess how individual symptoms may relate. In light of the TOPPITS results, a greater understanding of symptom reporting is required to generate further, much need research into appropriate therapies for patients with persistent throat symptoms.

Much has been written on the signs that may be related to reflux within the throat. The literature was summarised in an evidence based conference and subsequently published, highlighting the lack of evidence regarding throat signs, attributed to reflux, and citing the need for quality research in this area[52]. The most commonly cited study suggesting an association between symptoms and signs was conducted on only 40 patients and showed a Spearman's rank correlation of 0.86. TOPPITS will give far greater scope for assessing any relationship that may exist between throat symptoms and signs.

Figure 1-1 The potential overlapping of throat symptoms



1.5 Thesis Aims and Objectives

1.5.1 *Principal aim*

To explore persistent throat symptoms and signs, in order to identify patterns of specific complaints and produce a more clinically meaningful classification of patient presentations.

1.5.2 *Objectives*

1. To comment on the generalisability of the TOPPITS population, to general throat symptom populations, using descriptive analyses for the baseline demographic data, patient reported outcome measures (PROMs) and clinician rated throat appearances for patients.

2. Assess the relationship between the different PROMs and clinician rated throat appearances with one another.

3. Attempt to devise a clinically useful classification system for persistent throat symptoms, achieved using:

   Exploratory factor analysis of a) the three questionnaires' individual items combined together, and b) the individual questionnaires.

   Relevant cluster analysis techniques.

1.6 Thesis Layout

Four chapters will set out the work performed to address each objective:

Chapter 2 Baseline Demographic Data

Chapter 3 Relation between outcome measures

Chapter 4 Exploratory Factor Analysis

Chapter 5 Cluster Analysis

Each chapter will comprise an introduction, methods, results, discussion and conclusions.

Chapter 6 will present the overall thesis conclusions.

The Appendices contain additional material referenced within the thesis.

References

Chapter 2 Baseline Demographic Data

2.1 Introduction

The objective for Chapter 2 is to present descriptive analyses for the baseline demographic data, patient reported outcome measures (PROMs) and clinician rated throat appearances for patients recruited to the TOPPITS trial, and comment on the generalisability of the TOPPITS population to general ENT practice.

2.2 Methods

2.2.1 *TOPPITS Data management*

The proposed thesis was discussed within the TOPPITS trial management group (TMG). Initially, the work in this thesis was planned to fall under the remit of one of the trial's prespecified objectives, detailed within the trial's Statistical Analysis Plan (SAP). However, it was felt that this arrangement would limit the thesis objectives and that greater scope for exploring the rich baseline dataset would be achieved with a stand-alone study. A proportionate review application was therefore submitted to the research and ethics committee, with a favourable response obtained in May 2018 (Appendices - Research Ethics Committee Approval).

The TOPPITS baseline dataset was transferred into SPSS (Statistical Package for the Social Sciences) format by the data manager for the clinical trial (Ruth Wood). All patient identifiable information was removed other than the unique trial identification number.

The analysis set included all patients recruited into the TOPPITS trial. As this is baseline data only, protocol violation did not affect the data set. There was no requirement to perform an intention to treat analysis. Patients who were found to be ineligible post-randomisation and protocol violators were retained in the analysis for this thesis.

The data set included the patient unique trial identification number. It did not contain the randomisation group. Variables included in the data set were:

- Age at recruitment into the trial
- Gender
- Body mass index
- Smoking status
- Baseline Reflux Symptom Index (RSI) – individual item scores and total score

- Baseline Comprehensive Reflux Symptom Score (CReSS) - individual item scores and total score
- Baseline Laryngopharyngeal Reflux – Health related Quality of Life score (LPR-HRQL) - individual item scores and total score
- Baseline Reflux Finding Score (RFS) - individual item scores and total score

2.2.2 *Descriptive Analyses*

All analyses were performed using SPSS version 24.

Demographics

Descriptive statistics for age and body mass index (BMI) include the mean, standard deviation, median, interquartile range and range. Age and BMI are displayed graphically using histograms. Gender and smoking status are described using as frequencies and percentages.

Reflux Symptom Index

The Reflux Symptom Index (RSI)[53] is a nine-item self-administered questionnaire (see Appendices - The Reflux Symptom Index). Each item is scored on a Likert scale zero to five, giving a total score range of 0-45. Higher scores represent increasing severity of patient reported symptoms. There are suggestions that the "heartburn" item may influence much of the variation in scores of the RSI in previous studies. The RSI was also analysed without the heartburn item; termed RSI-HB, with a range of 0-40.

Descriptive statistics for both the RSI and the RSI-HB include the mean, standard deviation, median, inter-quartile range, and range for the total scores. The mean total score for the RSI is reported. However, for each item on the questionnaires (given the construct of each item), the median score and range are recorded in addition to the mean. The data was assessed for normality. The RSI and RSI-HB are represented graphically (histograms and normal curves). Individual items from the RSI are ranked according to median scores.

Data were assessed for normality using graphical methods. An approximate rule to assess normality is to consider the Kurtosis and Skewness scores of the data, with a Kurtosis value between -3 and +3 and a Skewness value between -0.8 and +0.8 suggesting normality.

Comprehensive Reflux Symptom Score

The Comprehensive Reflux Symptom Score (CReSS) [54] is a 34 item self-administered questionnaire of oesophageal and extra-oesophageal symptoms which has been tested on groups of "throat" patients, healthy controls and those attending for an upper gastrointestinal endoscopy (see Appendices - The Comprehensive Reflux Finding Score). Each item is scored on a zero to five Likert scale, giving a range of total scores 0-170. Descriptive statistics for the total score include the mean, standard deviation, median, inter-quartile range, and range. For individual factors the scores are presented as median and ranges. Means and standard deviations were derived for each item on the questionnaire and ranked. The data are also presented graphically.

Laryngopharyngeal Reflux Health Related Quality of Life Questionnaire

The Laryngopharyngeal Reflux Health Related Quality of Life questionnaire (LPR-HRQL) is a 43 item self-administered questionnaire[48] (Appendices - The Laryngopharyngeal Reflux Health Related Quality of Life Questionnaire). It is composed of four domains (Voice – 12 items, Cough – six items, Throat Clearing – six items, and Swallow – five items). Each of these items is scored on a zero to six Likert scale. The total for each domain is scored by taking the total score, subtracting the mean and dividing by the standard deviation – to give the z-score for each domain. Following each set of domain questions there is a "thermometer" question for that domain, scored 1 – 10. The last 10 questions cover the domain "overall impact of acid-reflux" and are scored 1-10. An "overall score" is calculated by adding the four thermometer scores to the 10 overall impact of acid reflux questions, to give a score of 14 – 140. The scoring instructions state that missing data can be imputed if less than six items are missing from the voice domain, or less than 3 for the cough, throat and swallow domains.

Descriptive statistics for the domain scores, thermometer scores and "overall impact of acid reflux" include the mean, standard deviation, median, inter-quartile range, and range. Medians and ranges were derived for each item on the questionnaire. The data are also presented graphically. The domain thermometer scores were ranked.

Reflux Finding Score

The Reflux Finding Score (RFS)[23] is a clinician assessed rating of the appearance of the larynx (Appendices - The Reflux Finding Score). There are eight items, with each item scored using varying categories. Some items score: zero = absent (clinical finding), two =present. Some items score: zero =none, two=mild, three=severe, four=obstructive. The total score range

is between 0-29. Descriptive statistics include the median, inter-quartile range, and range. Absent/present items were described according to frequencies and percentages. These are also assessed graphically.

2.3 Results

The TOPPITS baseline dataset contained data on 344 individual patients, recruited over 8 centres in the UK. Institutional recruitment was as follows: Newcastle – 132 (38%), Nottingham – 70 (20%), Sunderland – 46 (13%), Glasgow – 39 (11%), Manchester – 27 (7.8%), Stockport – 11 (3.2%), Birmingham – 10 (2.9%), Brighton – 9 (2.6%).

Initial data screening revealed five patients with duplicate identification numbers and no demographic or questionnaire data. These were removed. Three patients were randomised to the TOPPITS trial but subsequently were found to be ineligible to continue in the trial. Baseline data existed for two of these patients and has been included in this thesis analysis. Data is present on 344 patients for this baseline analysis.

There was no missing data for patients' age or gender. Data was missing on Body Mass Index for five patients and on smoking status for five patients. Patients with missing data for these demographics were excluded from the descriptive statistical analysis.

2.3.1 *Patient Demographics*

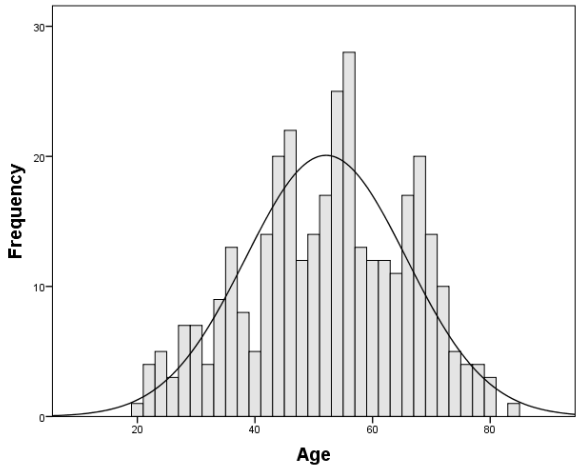
Over half of the population were female (195, 56.7% - 149 males, 43.3%) with mean age 52.1 years (sd 13.7, Table 2-1). The ages were normally distributed (Figure 2-1 shows low Kurtosis and Skewness values; Kurtosis -0.59. Skewness -0.20). The mean BMI was 28.1 (Figure 2-1) with missing data for five patients. The BMI scores were not normally distributed, given a Kurtosis score outside of an acceptable range of -3 to +3, and a Skewness score outside an acceptable range of -0.8 to +0.8 (Figure 2-2- Kurtosis 3.28 . Skewness 1.28). Figure 2-1 also shows a box and whisker plot of the BMI scores. The box depicts the mean and interquartile range with the whiskers showing the range, with separate outlying scores. There were two particular outliers with BMIs of 11.3 and 56.9. It is possible that these were errors in the data collection at source, given how extreme the two values are. With these two outliers removed the BMI distribution is positively skewed, as expected if compared to the general UK population (Kurtosis = 1.83. Skewness = 0.97 – which remains greater than 0.8). There were 100 patients who were smokers (29.5%) with missing data for 5 patients. Amongst smokers, , with a median cigarette consumption was 10 pack years, with the distribution of pack years smoked being positively skewed (Table 2-1 and Figure 2-1).

Table 2-1 Baseline distribution of demographics

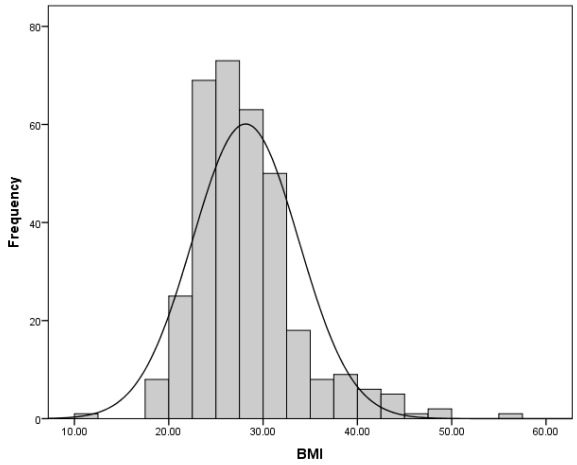
Demographic (n=)	Mean	SD	Median	IQR	Range
Age (344)	52.1	13.7	53	20 (43 – 63)	20 – 84
BMI (339)	28.1	5.62	27.1	6.05 (24.5 – 30.6)	11.3 – 56.9
"Pack years smoked" if current smoker (100)	14.7	13.8	10	15 (5 – 20)	0.5 - 76

Figure 2-1 Graphical representations of Age and BMI distributions

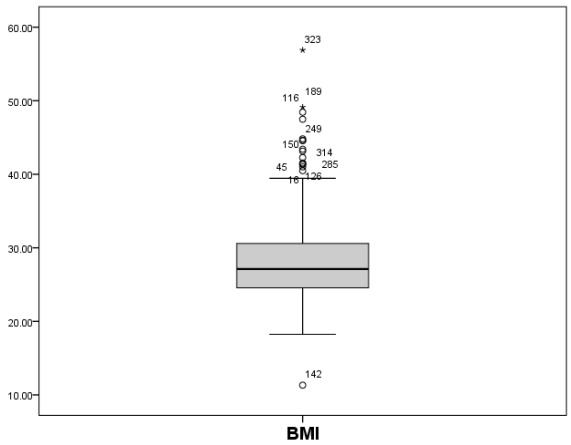
Histogram showing age distribution with overlying normal curve



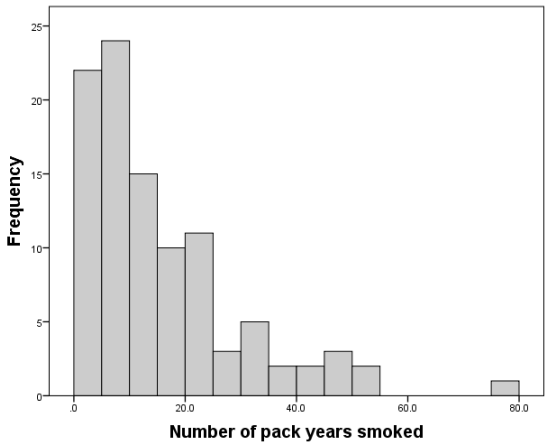
Histogram showing BMI distribution with overlying normal curve



Box and whisker plot of BMI distribution



Histogram showing distribution of "pack years smoked" in current smokers



2.3.2 *Questionnaire Data*

2.3.2.1 Reflux Symptom Index (RSI)

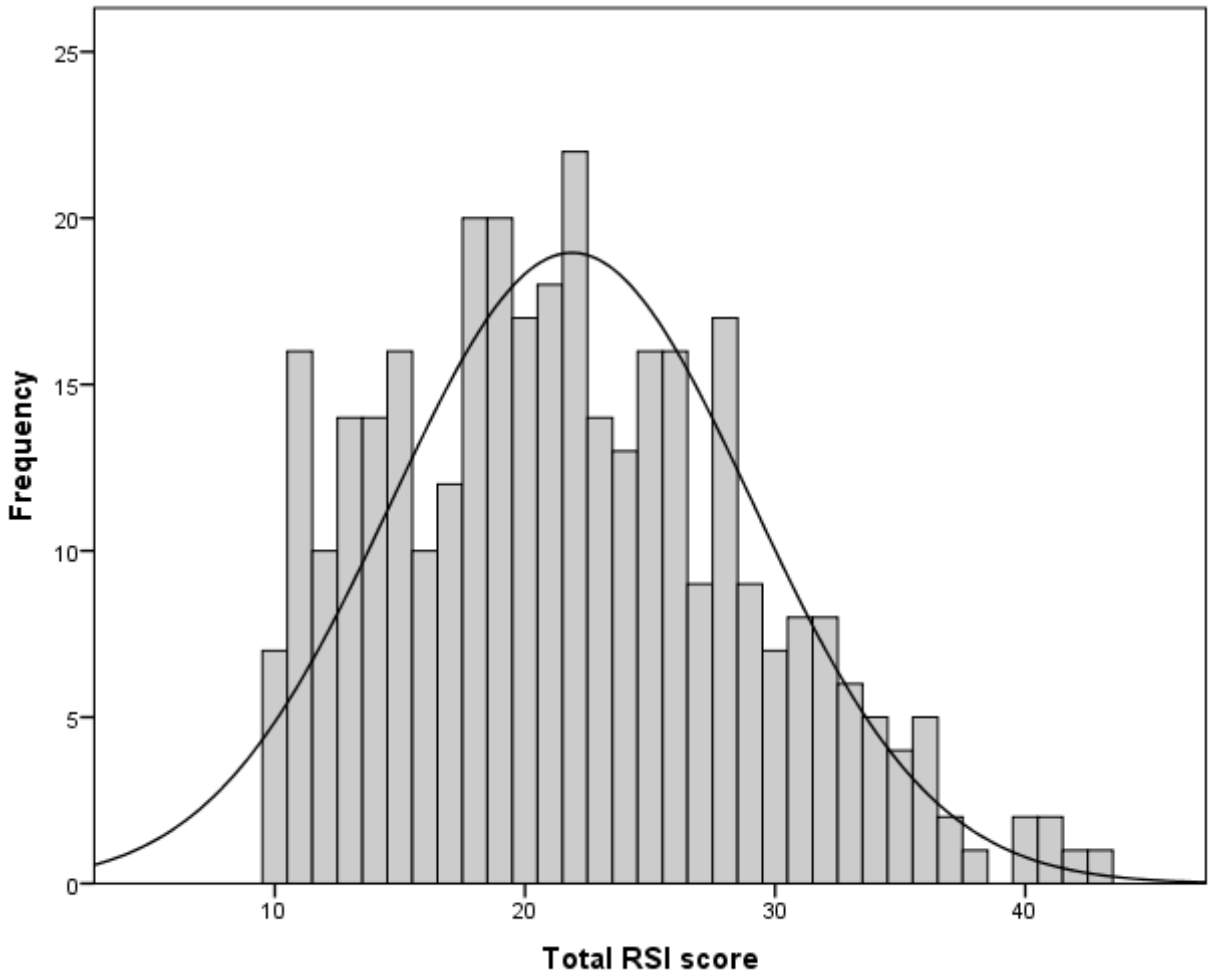
The total score RSI data was missing for two patients; both with single items missing from their questionnaires. The mean total RSI score was 21.9 (sd. 7.2, Table 2-2) and was normally distributed (Figure 2-5). Table 2-3 and Table 2-4 show the scores for the individual RSI item scores. The items "sensation of something caught in the throat or lump in the throat" and "throat clearing" had the highest population mean and median. "Difficulty swallowing

food, liquids or tablets" and " breathing difficulties or choking episodes" had the lowest population mean scores. Figure 2-3 shows the graphical representation of the RSI items ranked according to population mean and median scores. Graphically, the individual RSI item scores were not normally distributed (Shown in the appendix Figure 7-1, Figure 7-2, Figure 7-3, Figure 7-5, Figure 7-6, Figure 7-7, Figure 7-8, Figure 7-9).

Table 2-2 Distribution of Total RSI scores

n=	342
Mean	21.9
Standard deviation	7.20
Median	21
Interquartile range	11 (16 – 27)
Range	10 – 43

Figure 2-2 Histogram showing distribution of Total RSI score with overlying normal curve



Kurtosis = -0.33. Skewness = 0.43

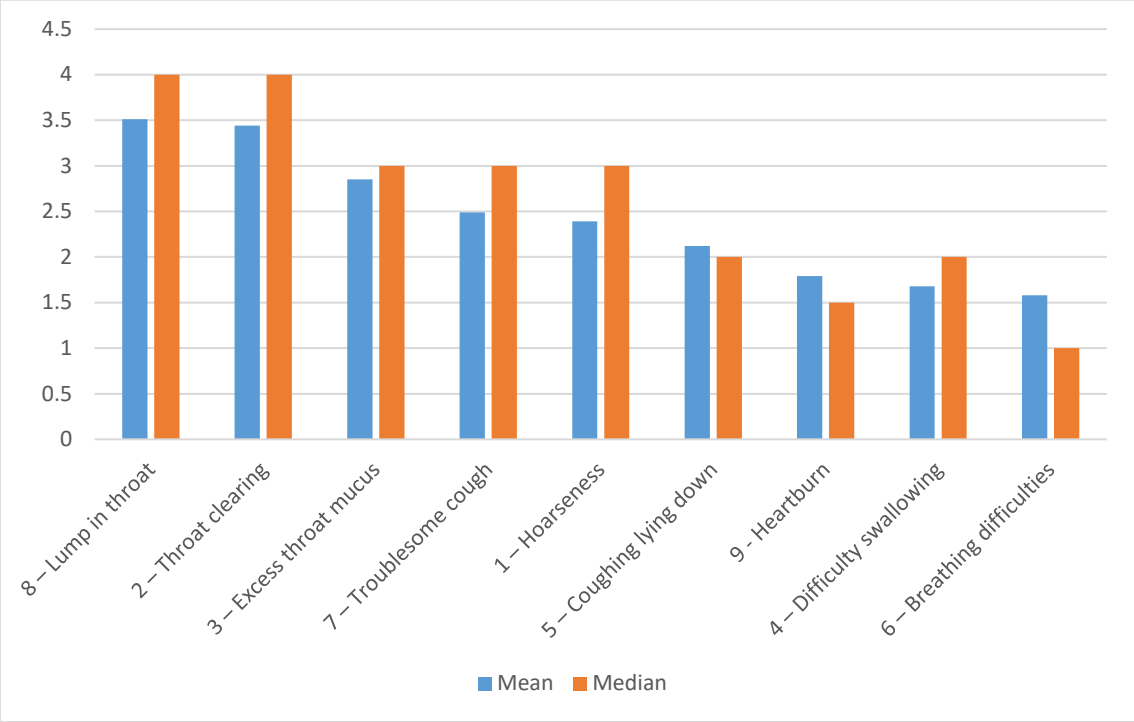
Table 2-3 Distribution of individual RSI items

RSI item	Mean	Standard deviation	Median	Inter-quartile range	Range
1 Hoarseness	2.39	1.62	3	3 (1 – 4)	0 - 5
2 Throat clearing	3.44	1.29	4	1 (3 – 4)	0 – 5
3 Excess throat mucus	2.85	1.56	3	2 (2 – 4)	0 - 5
4 Difficulty swallowing	1.68	1.61	2	3 (0 – 3)	0 – 5
5 Coughing lying down	2.12	1.64	2	3 (0 – 3)	0 – 5
6 Breathing difficulties	1.58	1.62	1	3 (0 – 3)	0 – 5
7 Troublesome cough	2.49	1.71	3	3 (1 – 4)	0 – 5
8 Lump in throat	3.51	1.44	4	2 (3 – 5)	0 – 5
9 Heartburn	1.79	1.60	1.5	3 (0 – 3)	0 – 5

Table 2-4 RSI items ordered by rank mean score

Rank mean score	RSI item	Mean score
1	8 Lump in throat	3.51
2	2 Throat clearing	3.44
3	3 Excess throat mucus	2.85
4	7 Troublesome cough	2.49
5	1 Hoarseness	2.39
6	5 Coughing lying down	2.12
7	9 Heartburn	1.79
8	4 Difficulty swallowing	1.68
9	6 Breathing difficulties	1.58

Figure 2-3 Bar chart showing ranked mean and median RSI item scores



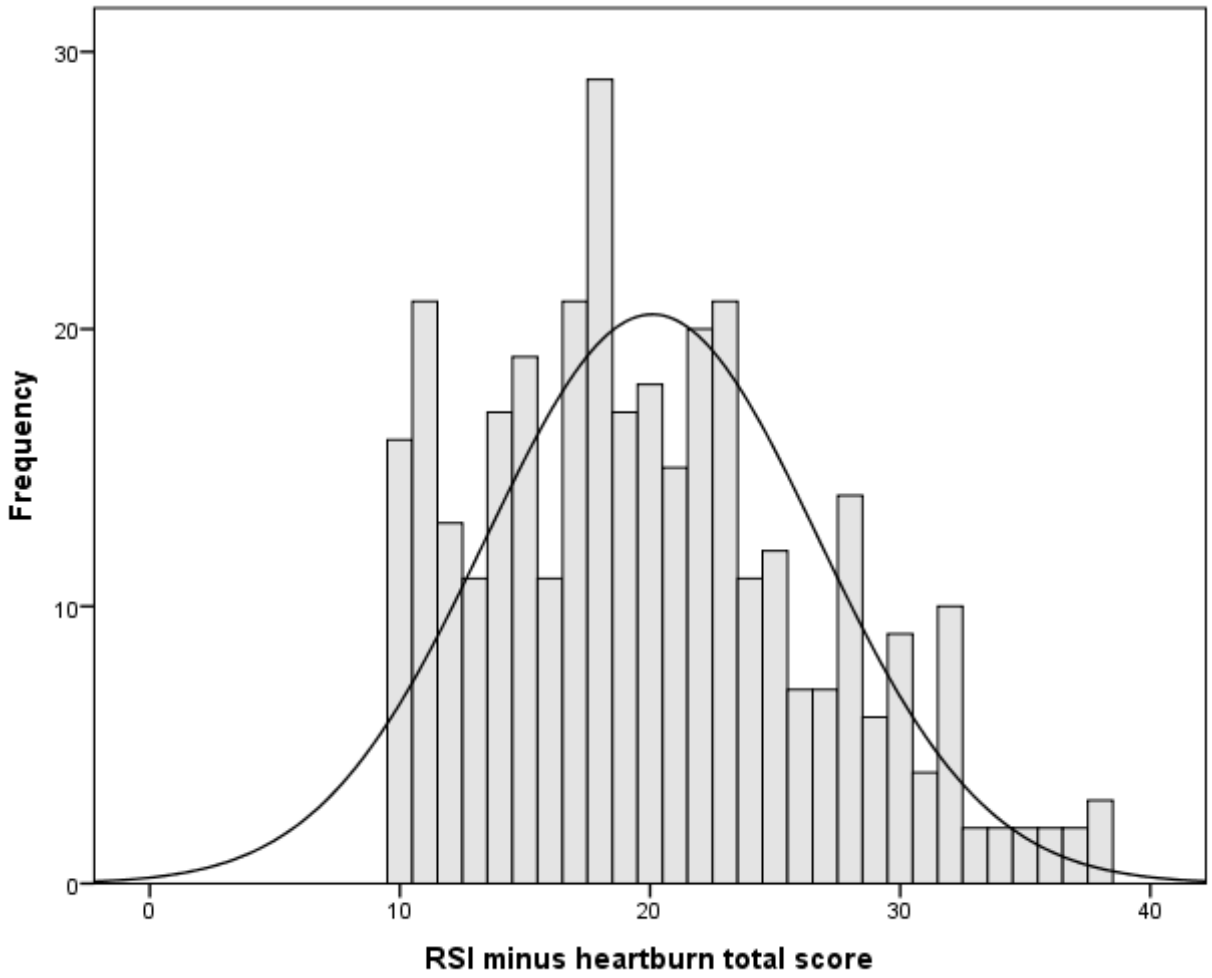
Reflux Symptom Index minus Heartburn score (RSI-HB)

With the ninth RSI item omitted, the mean RSI –HB score was 20.1 (Table 2-5). The distribution of the RSI-HB scores appeared normal (Figure 2-4).

Table 2-5 Distribution of Total RSI-HB scores

n=	342
Mean	20.1
Standard deviation	6.65
Median	19
Interquartile range	9 (15 – 24)
Range	10 – 38

Figure 2-4 Histogram showing distribution of Total RSI-HB scores with overlying normal curve



Kurtosis = -0.37. Skewness = 0.49

2.3.2.2 Comprehensive Reflux Symptom Score

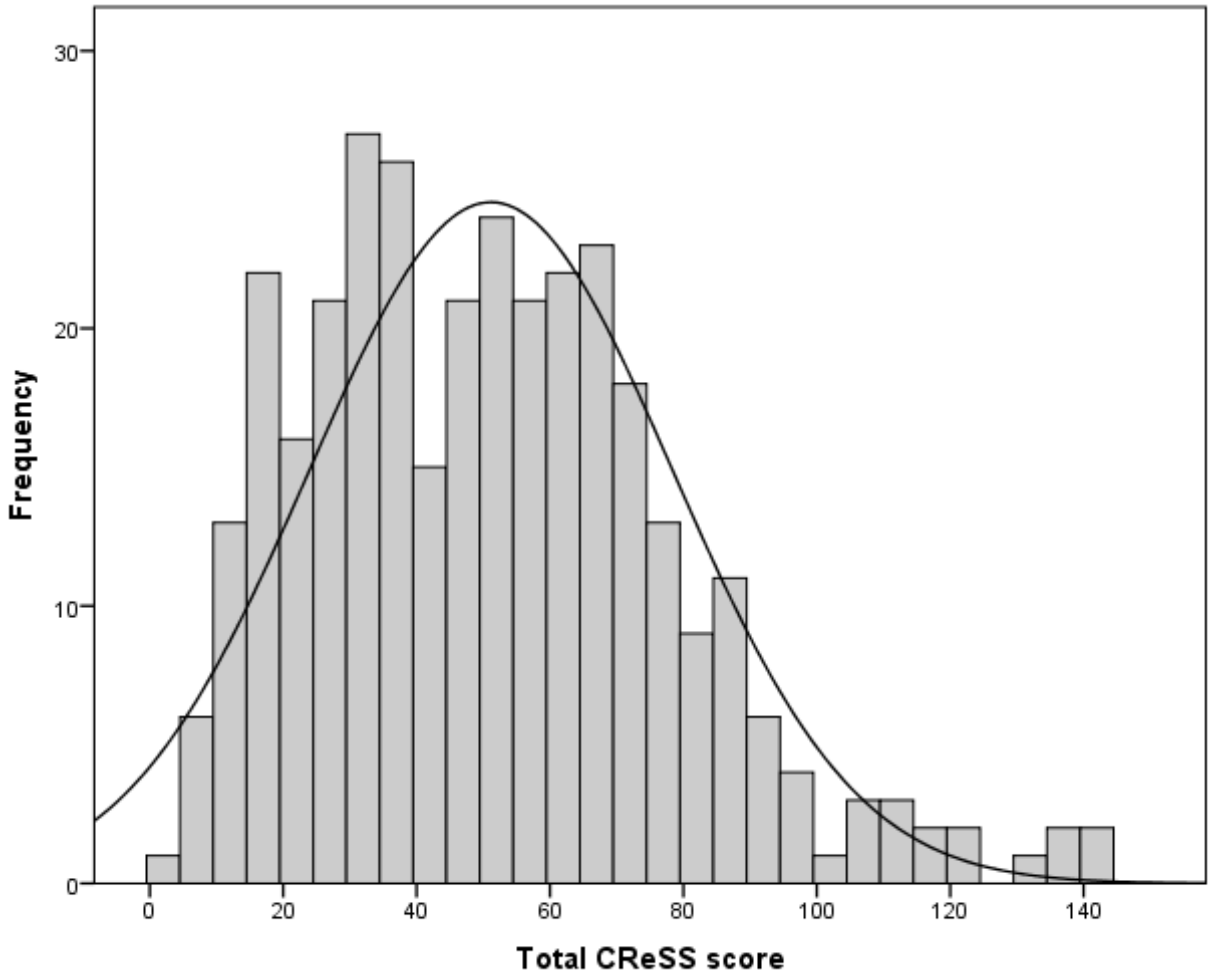
Data was missing for nine patients: five with no CReSS submitted and four with several missing items. These nine patients were removed from the analysis.

The mean total score CReSS score was 51.2 (s.d. 27.2, Table 2-6). The total CReSS score appeared normally distributed, with Kurtosis and Skewness scores within an acceptable range (Figure 2-5). Table 2-7 and Table 2-8 show the descriptive data and ranking of the individual CReSS items. The highest ranked items were “throat clearing”, “feeling things stuck in throat” and “lump in throat”. The lowest ranked items were “vomiting”, “hiccups” and “decreased appetite”.

Table 2-6 Distribution of Total CReSS scores

n=	335
Mean	51.2
Standard deviation	27.2
Median	49
Interquartile range	39 (30 - 69)
Range	2 - 142

Figure 2-5 Histogram showing distribution of Total CReSS score with overlying normal curve



Kurtosis = 0.27. Skewness = 0.71

Table 2-7 Descriptive data for the individual CReSS items scores

CReSS Item	Mean	SD	Median	Interqu-artile range	Range
1 Heartburn	1.36	1.46	1	2 (0 – 2)	0 – 5
2 Pressure in chest	1.09	1.42	0	2 (0 – 2)	0 – 5
3 Regurgitation	1.12	1.36	1	2 (0 – 2)	0 – 5
4 Acid/sour taste in mouth	1.50	1.48	1	3 (0 – 3)	0 – 5
5 Gurgling stomach	1.69	1.56	1	3 (0 – 3)	0 – 5
6 Lump in throat	2.86	1.74	3	3 (1 – 4)	0 – 5
7 Difficulty swallowing food	1.44	1.61	1	3 (0 – 3)	0 – 5
8 Difficulty swallowing liquids	0.83	1.23	0	1 (0 – 1)	0 – 5
9 Nausea	0.91	1.38	0	2 (0 – 2)	0 – 5
10 Pain in throat	1.49	1.60	1	3 (0 – 3)	0 – 5
11 Vomiting	0.42	0.99	0	0 (0 – 0)	0 – 5
12 Bloating	1.36	1.63	1	2 (0 – 2)	0 – 5
13 Belching	1.42	1.57	1	3 (0 – 3)	0 – 5
14 Flatulence	1.52	1.57	1	3 (0 – 3)	0 – 5
15 Hiccups	0.62	1.07	0	1 (0 – 1)	0 – 5
16 Decreased appetite	0.73	1.28	0	1 (0 – 1)	0 – 5
17 Rush of saliva in mouth	1.20	1.47	1	2 (0 – 2)	0 – 5
18 Feeling full early	1.14	1.49	0	2 (0 – 2)	0 – 5
19 Bad breath	1.05	1.41	0	2 (0 – 2)	0 – 5
20 Back pain	1.92	1.74	2	3 (0 – 3)	0 – 5
21 Headache	1.38	1.47	1	2 (0 – 2)	0 – 5
22 Choking	1.14	1.53	0	2 (0 – 2)	0 – 5
23 Coughing upright	1.99	1.61	2	3 (0 – 3)	0 – 5
24 Coughing after eating	1.49	1.56	1	3 (0 – 3)	0 – 5
25 Coughing lying down	1.95	1.73	2	3 (0 – 3)	0 – 5
26 Wheezing	1.07	1.52	0	2 (0 – 2)	0 – 5
27 Difficulty breathing	1.10	1.57	0	2 (0 – 2)	0 – 5

CReSS Item	Mean	SD	Median	Interqu-artile range	Range
28 Hoarseness	2.08	1.76	2	4 (0 – 4)	0 – 5
29 Throat clearing	3.20	1.50	4	2 (2 – 4)	0 – 5
30 Excess mucus	2.41	1.82	2	3 (1 – 4)	0 – 5
31 Mucus dripping in throat	2.06	1.89	2	4 (0 – 4)	0 – 5
32 Feeling things stuck in throat	2.89	1.79	3	3 (1 – 4)	0 – 5
33 Indigestion	1.30	1.45	1	2 (0 – 2)	0 – 5
34 Stomach acid coming up	1.41	1.55	1	3 (0 – 3)	0 – 5

Table 2-8 CReSS items ordered by rank mean score

Rank mean score	CReSS Item	Mean Score
1	29 Throat clearing	3.20
2	32 Feeling things stuck in throat	2.89
3	6 Lump in throat	2.86
4	30 Excess mucus	2.41
5	28 Hoarseness	2.08
6	31 Mucus dripping in throat	2..06
7	23 Coughing upright	1.99
8	25 Coughing lying down	1.95
9	20 Back pain	1.92
10	5 Gurgling stomach	1.69
11	14 Flatulence	1.52
12	4 Acid/sour taste in mouth	1.50
13	10 Pain in throat	1.49
14	24 Coughing after earing	1.49
15	7 Difficulty swallowing food	1.44
16	13 Belching	1.42
17	34 Stomach acid coming up	1.41
18	21 Headache	1.38

Rank mean score	CReSS Item	Mean Score
19	1 Heartburn	1.36
20	12 Bloating	1.36
21	33 Indigestion	1.30
22	17 Rush of saliva in mouth	1.20
23	18 Feeling full early	1.14
24	22 Choking	1.14
25	3 Regurgitation	1.12
26	27 Difficulty breathing	1.10
27	2 Pressure in chest	1.09
28	26 Wheezing	1.07
29	19 Bad breath	1.05
30	9 Nausea	0.91
31	8 Difficulty swallowing liquids	0.83
32	16 Decreased appetite	0.73
33	15 Hiccups	0.62
34	11 Vomiting	0.42

2.3.2.3 Laryngopharyngeal Health-Related Quality of Life (LPR-HRQL)

There was missing data for 10 patients from the Voice domain, 12 patients from the Cough domain, 10 patients from the Throat domain, and six patients from the Swallow domain. Following imputation of missing individual items across the domains, according to the pre-defined methodology, there was missing data for six patients from each of the four domains. These patients have been removed from the analysis.

Table 2-9, Table 2-10 and Table 2-11 show the descriptive data for the domain scores, thermometer scores and “overall impact of acid” scores within the LPR-HRQL. Given the varying number of items within the domains, Table 2-9 gives a mean score per item for each domain. This shows that the throat clearing domain had the highest mean score per item, with voice the lowest score.

Table 2-9 Descriptive data for the domain scores in the LPR-HRQL

Domain	n =	Mean	SD	Median	Interquartile range	Range	Mean / no. of items	Rank
Voice	338	14.9	15.2	8	14 (6 – 20)	0 – 67	1.24	4
Cough	338	8.56	9.06	5	13.25 (0.75 – 13)	0 – 35	1.43	2
Throat clearing	338	9.27	7.38	8	11 (3 – 14)	0 – 35	1.55	1
Swallow	338	7.12	6.69	5	9 (2 – 11)	0 - 30	1.42	3

Table 2-10 Descriptive data for the thermometer scores in the LPR-HRQL

Domain	n =	Mean	SD	Median	Interquartile range	Range	Rank
Voice	338	3.33	2.81	2	4 (1 – 5)	0 – 10	4
Cough	338	3.54	2.68	3	4 (1 – 5)	0 – 10	3
Throat clearing	338	4.18	2.67	4	4 (2 – 6)	0 – 10	2
Swallow	338	4.35	2.74	4	5 (2 – 7)	0 – 10	1

Table 2-11 Descriptive data for the overall score for the LPR-HRQL

n =	338
Mean	50.8
Standard deviation	28.0
Median	46
Interquartile range	44 (26 – 70)
Range	14 - 134

2.3.3 *Reflux Finding Score (RFS)*

There was missing data for 90 patients. RFS scores were not obtained for these patients. Explanations for this missing data are covered in the discussion. Table 2-12, Table 2-13 and Figure 2-6 show the descriptive data and distribution of the total RFS scores and individual RFS items. The mean score was 8.78 (s.d. 4.14). Given the variable scoring range for the individual RFS items, no one item was particularly higher than the others. Whilst the erythema score was the highest mean score, this item had a larger standard deviation and interquartile range than posterior commissure oedema.

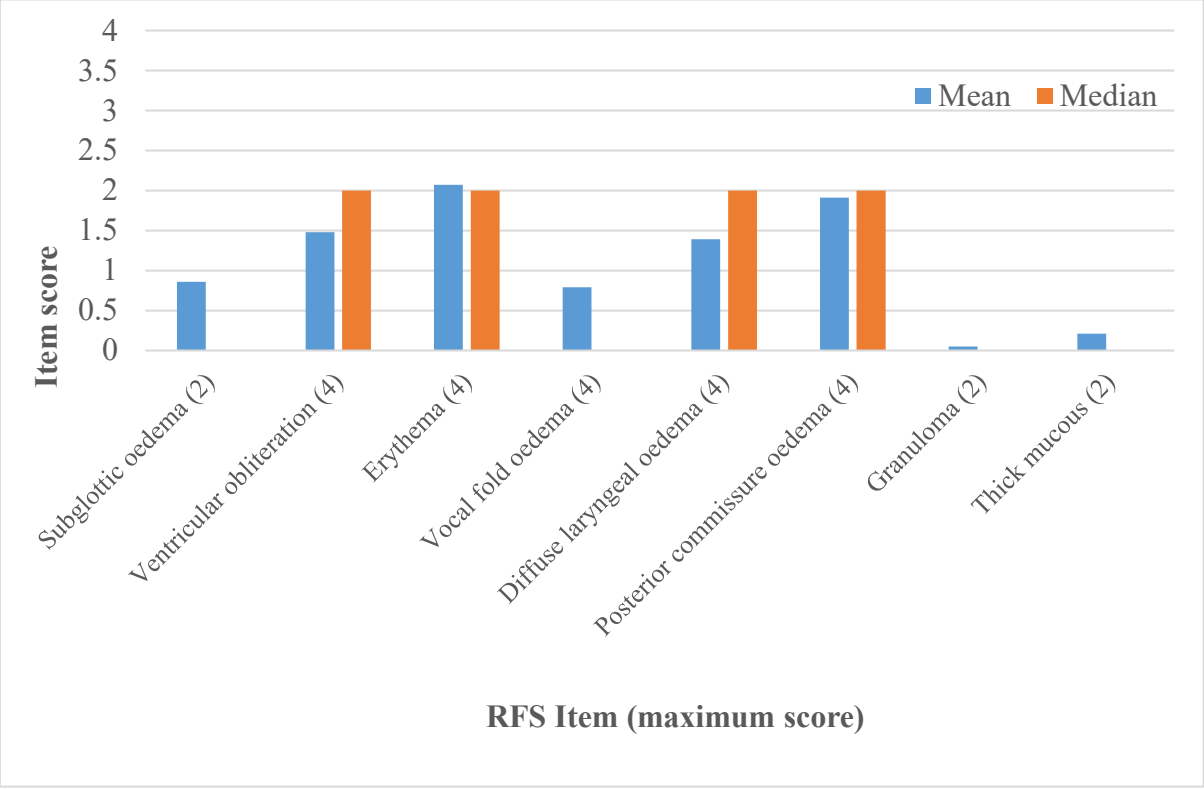
Table 2-12 Descriptive data for the Total RFS scores

n =	254
Mean	8.78
Standard deviation	4.14
Median	9
Interquartile range	5 (6 – 11)
Range	0 - 24

Table 2-13 Descriptive data for the individual RFS item scores

RFS Item (maximum score)	Mean	SD	Median	Interquartile range	Mode
Subglottic oedema (2)	0.86	0.99	0	2 (0 – 2)	0
Ventricular obliteration (4)	1.48	1.25	2	2 (0 – 2)	2
Erythema (4)	2.07	1.51	2	4 (0 – 4)	2
Vocal fold oedema (4)	0.79	1.00	0	2 (0 – 2)	2
Diffuse laryngeal oedema (4)	1.39	0.99	2	2 (0 – 2)	2
Posterior commissure oedema (4)	1.91	0.90	2	2 (1 – 3)	2
Granuloma (2)	0.05	0.31	0	0	0
Thick mucus (2)	0.21	0.62	0	0	0

Figure 2-6 Histogram to show distribution of RFS individual items



2.4 Discussion

2.4.1 *Demographics*

In common with many other reported studies assessing throat symptoms, there were more females (57%) than males initially recruited to the TOPPITS trial. A case series of 1044 patients with similar symptoms had an identical proportion of females[28]. Within the TOPPITS population, the mean age of 52 and BMI of 28 are very similar to those found in other interventional trials of throat symptom management[55].

2.4.2 *RSI comparison with published data*

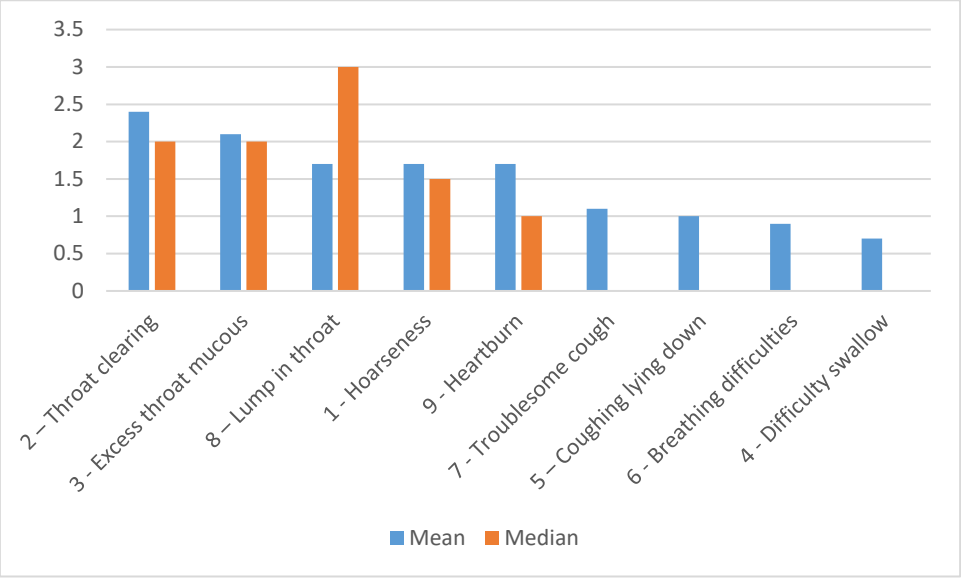
The RSI was chosen as the primary outcome measure in the TOPPITS trial on the basis of its popularity in the published literature, allowing comparisons to be made and allowing the trial's results to be reflected into common practice. The RSI mean was 21.9 (sd 7.2). Accepting that the minimum score in the sample has to be 10, given the eligibility criteria for TOPPITS, graphically and with low kurtosis and skewness scores, the data can be considered normally distributed. The original study that described the RSI on 25 symptomatic patients found the RSI to be 20.9 (9.6) in one of the baseline measurements. The largest case series published quote a mean pre-treatment RSI values of 15.1 in a series of 455 patients [27] and 13.8 in a series of 1044 patients[28]. Lee et al [27] do not quote a standard deviation for the entire group, but do so for the individual RSI items and also for subgroups. For example the RSI for patients aged between 30 and 40 was 16.2 (sd 8.1). These figures are lower than the baseline values for the RSI in the TOPPITS population, but the cited publications are from all comers presenting to a specialist clinic, and not those patients eligible for a clinical trial. In comparison with other interventional trials, Reichel et al[56] recruited 62 patients with an RSI >13 (TOPPITS stated RSI without the heartburn item >10); the mean RSI in the group receiving PPI was 23.1 (sd 7.5) and in the control group 21.8 (sd 6.7). McGlashan et al recruited 49 patients with an RSI > 10 to a randomised controlled study assessing a liquid alginate; the RSI in the treatment group was 23.9 (sd 7.0) and in the control group 24.6 (7.4). In a recent non-randomised study of the effects of alginate in addition to PPI in 72 respondents, the RSI was 19.2 (sd 7.4 – calculated using the published confidence intervals) in the alginate group and 21.3 (9.0) in the alginate and PPI group. Whilst the mean RSI in the published populations varies, with lower scores in non-selected case series and higher scores in trials with entry criteria, an interesting finding throughout is the similarity in standard deviation between trials.

During the design phase of the TOPPITS trial, the research team were concerned that PPI treatment would produce significant improvements in the traditional heartburn symptom, over

placebo. During the evidence based medicine conference in 2011, this concern was discussed and many clinicians were sceptical that the improvements in the RSI scores in case series of patients treated with PPIs may have been largely due to improvements in heartburn, for which PPIs are effective. For this reason the eligibility criteria for TOPPITS was based on a symptom score omitting the RSI heartburn item. The results presented by Reichel et al[56] would support this concern over preferential heartburn improvements with PPI. Whilst they concluded that symptoms improved in 30 patients receiving PPI compared to 28 receiving placebo, as measured by a change in total RSI, the only single symptom that was significantly different between the two groups was the heartburn item of the RSI.

Selecting patients on the basis of the RSI without the heartburn item for entry into the TOPPITS trial could have resulted in a trial with less patients with traditional heartburn as their principle symptom than would otherwise normally be seen in clinics. Ranking the RSI individual items has shown that the heartburn item was the seventh highest score. Figure 2-7 shows the RSI items ranked, from the largest published series of 1044 patients[28]. The heartburn item was ranked fifth.

Figure 2-7 Histogram to show ranked RSI items from Haberman et al.



Lee et al similarly found that the heartburn item was ranked fifth on the RSI[27]. The highest scored items in the TOPPITS baseline data are consistent with those from the two comparative large series; a feeling of a lump in the throat, throat clearing and excess mucus. The comparative studies may have found the heartburn item slightly higher ranked than in TOPPITS, but the throat symptoms reported by patients are consistent. The baseline data from

TOPPITS would therefore appear generalisable and representative of patients presenting to standard ENT clinics.

The clinical relationship between heartburn and persistent throat symptoms is a matter of debate and requires further research to define it. Whilst it is argued that LPR and GORD are separate entities, clinicians must appreciate that traditional GORD, or heartburn symptoms are prevalent in up to a fifth of the general population. One-fifth of patients presenting with persistent throat symptoms may therefore report co-existing GORD symptoms.

2.4.3 *CReSS comparison with published data*

There are no other interventional trials that have reported on CReSS outcomes. Drinnan et al[45] evaluated the CReSS with 639 patients surveyed. The mean total CReSS score for 177 patients presenting to ENT with persistent throat symptoms was 32.3 (sd 21.8), for 136 patients undergoing gastroenterology delivered endoscopy for proven oesophagitis the total score was 44.2 (26.9), for 223 patients undergoing endoscopy for non-oesophagitis the total score was 39.6 (28.7), and for 100 controls the total score was 6.1 (7.5). The ENT group had the highest percentage of positive responders to throat clearing (80%), hoarse voice, lump and excess mucus. The oesophagitis group had the highest percentage with heartburn (82%) and indigestion. Those undergoing oesophagogastroduodenoscopy without oesophagitis had slightly lower scores than those patients with oesophagitis, but the heartburn (72%) and indigestion items remained the highest reported in this patient group.

The mean total CReSS score in the baseline TOPPITS data was 51.2 (27.2); higher than any of the groups reported by Drinnin et al. This may reflect the entry criteria into TOPPITs, selecting patients with higher symptom reporting.

2.4.4 *LPR-HRQL comparison with published data*

This appears a complex questionnaire, not necessarily for patients to complete, but in terms of the appropriate method to analyse the results. The LPR-HRQL has not been readily adopted. Its validity was thoroughly assessed when reported by 117 patients receiving PPIs for “LPR”; Carrau at el reported the domain scores of voice, cough, throat and swallow along with the overall impact of acid score[57]. No mention is made of how to report the thermometer scores that accompany each domain section. Andersson et al reported on the use of the LPR-HRQL in validating a Swedish Pharyngeal Reflux Symptom Questionnaire, but no values were published for the LPR-HRQL specifically[58]. Vaezi et al[55] did report the total LPR-HRQL scores in 145 patients enrolled in an RCT to assess PPI effectiveness. Unfortunately, it is not clear from the methods which score they presented; whether it was a score of 140 or whether

they had converted it into a score out of 100. The baseline scores were: mean (sd) for the treatment group 30.9 (20.6), and for the placebo group 27.9 (19.3). Lee et al do reported the LPR HRQL outcomes in 180 patients treated with PPIs[59] with a mean RSI of 13.2 (sd 8.7). Similar to Carrau et al, the domain scores and overall acid scores are presented. The total score and the thermometer scores were not reported. Table 2-14 shows how the TOPPITS scores compare to the two published series.

Table 2-14 Comparison with published LPR-HRQL domains

LPR-HRQL domain	Carrau et al[57] mean (SD)	Lee et al[59] mean (SD)	TOPPITS mean (SD)
Voice	19.7 (16.1)	18.7 (20.7)	14.9 (15.2)
Cough	7.6 (8.8)	8.8 (10.5)	8.6 (9.1)
Clear Throat	9.6 (7.6)	7.4(8.7)	9.3 (7.4)
Swallow	7.6 (5.8)	7.3 (7.3)	7.1 (6.7)
Overall impact of acid	32.1 (20.9)	23.7 (17.5)	50.8 (28.0)

The individual domains of the questionnaire are problematic to compare as the voice domain has 12 items, the cough domain 6, throat 6 and swallow 5. Given the differing number of items in the domains, this analysis calculated the mean score per item within each domain. The throat clearing domain had the highest score, which would be in keeping with the symptoms ranked highest in the RSI and CReSS. However, when the thermometer scores were ranked, the swallow domain was ranked first with the throat thermometer score ranked as third. A lack of concordance between mean domain item scores and the domain thermometer scores does raise the question as to whether the summary thermometer score truly represents the domain items. Interestingly, the TOPPITs population appears to have a higher overall impact of acid, compared to the other two studies, despite similar scores in the other domains. The domain questions ask specific symptom questions. The domain thermometers and the overall impact of acid questions ask how particular symptoms affect "quality of life". For Chapter 4 and Chapter 5, only the 4 domains items and scores will be used in the exploratory factor analysis and cluster analysis.

RFS

The mean RFS score for the baseline TOPPITS data was 8.78 (sd 4.1). Interestingly this is lower than the group of 220 patients who were included in the final TOPPITS primary outcome analysis. When the patients who had completed the primary outcome measure within the specified window were included, the baseline RFS for this group was 9.4 (3.9).

Even with the use of a specified scoring tool, assessing the appearance of the throat in a standardised manner is demanding. The quality of the endoscope used or the printed / saved image will affect appearances. Subjectivity and reporter bias are clearly problems. TOPPITs used digital images saved as high-quality data files. These were then analysed by an independent clinician, who had no access to the patients' symptom scores. There was a significant volume of missing data for the RFS. When discussed in the TOPPITS trial management group, the data tended to be missing for whole groups and time periods of recruited patients. Hence, the missing data appeared to be a problem with capturing, storing and transferring the images from recruiting centres rather than there being any suggestion of selection bias.

The published RFS scores do vary. Haberman et al quoted a mean RFS of 15.8 in 1044 patients with a mean RSI of 13.8[28]. This seems a particularly high RFS given the patient reported symptoms on the RSI were a lot lower than in the TOPPITS population. Lee et al stated the mean RFS to be 9.1 in 455 patients with a mean RSI of 15.1[27]. Interventional trials with a minimum RFS entry criteria quote up to 14.9 (sd 2.5) and 10.4 (3.6) [32, 56].

Figure 2-6 shows that if the scoring system for each RFS item is taken into account, no single item stands out in the TOPPITS baseline population as being particularly frequently reported. Granuloma formation and thick mucus appear infrequently reported which would be consistent with the author's experience. Granuloma formation is a particular laryngeal finding, often associated with pain. They are so infrequently found that demonstrating any causal effect by reflux, a commonly held theory, would seem impossible to achieve.

As no individual item stands out on the RFS, it would appear sensible to report the RFS as a total score and not by individual items.

2.5 Conclusions

The patient demographics and outcome measures scores from the TOPPITs baseline data are comparable to the published literature. The highest scored symptoms were those of a lump in the throat, throat clearing and excess throat mucus. Traditional heartburn symptoms were not ranked highly and this finding is again in keeping with the published literature. The baseline TOPPITS data would therefore appear generalisable and the results from the subsequent chapters of this thesis can be applied to other patients presenting with persistent throat symptoms.

Chapter 3 Relation between outcome measures

3.1 Introduction

Thus far, the thesis has demonstrated that the baseline profile of patients recruited to the TOPPITS trial is comparable to similar published interventional trials, in terms of demographics, symptoms and throat appearances. TOPPITS offers the unique opportunity to assess how three patient reported outcome measures compare and to assess whether the symptom reporting relates to throat appearances.

TOPPITS collected patient reported symptoms using three questionnaires: the RSI, the CReSS, and the LPR HRQL. The background to the questionnaires is detailed in Chapter 1. The RSI appeared the most popular symptom tool in terms of its frequency of use in published studies, but the TOPPITS research team had reservations regarding the ninth polysymptomatic item of heartburn, indigestion, chest pain and acid coming up. The CReSS was included as an expanded 34 item questionnaire, covering all the RSI items but in individual questions. Whilst not a specific objective of TOPPITS, both the RSI and CReSS were completed by patients to allow future analysis and assess if one is superior in applicability. The LPR-HRQL was used in TOPPITS as it seemingly offered unique quality of life data. Whilst it would be expected that three questionnaires assessing the same types of symptoms should be positively related in scoring profiles, this is the first study to formally assess the relationship between these outcome measures.

The relationship between persistent throat symptoms and throat appearances is under debate. If the described entity of "LPR" truly exists, and it is accepted that the diagnosis of such a condition should rely on symptoms, signs and objective testing, then a relationship should be demonstrable between these factors. Interventional trials have set inclusion criteria based on minimum presenting symptom scores (RSI) and signs (RFS), implying a perceived relationship[32, 56]. Given the large number of observational cohort studies reporting the RSI and the RFS it is quite interesting to see how few have analysed the relationship between the two measures. The single frequently reported study was performed by Mesallam et al[26]. Forty randomly selected voice clinic attendees were retrospectively analysed with their presenting RSI and recorded laryngeal endoscopy images. Six clinicians, which reduced to four (one removed due to poor inter-rater reliability – which could be viewed as unacceptably altering the set methodology, and one removed as they could not complete the second evaluation), scored the RFS to assess inter-rater reliability (the kappa statistic ranged between +0.59 and 0.79). The RSI mean was 20.2 (sd 4.6) and the RFS mean 9.3 (4.4); comparable

with the TOPPITS baseline data. The Spearman's rank coefficient of correlation between the RSI and RFS was r= 0.86 ($p< 0.0001$).

Age may impact on symptom reporting. Lechien et al found that the baseline RSI was lower in older patients, within a cohort of 80 patients treated with PPIs and divided into three age groups [60]. However, their age did not appear to affect their reported response to therapy. Lee et al also showed that presenting RSI scores reduced with increasing age[27].

Body Mass Index would intuitively appear related to reflux. Jacobson et al demonstrated a dose dependent relationship between increasing BMI and frequent reflux symptoms, among 10,545 women [61]. Hence, if LPR is a genuine entity with a name suggesting a link to reflux, one would anticipate a relationship between increasing BMI and reported throat symptoms. The evidence is not clear on BMI effects on throat symptoms. Sone et al administered the RSI to 410 attendees of a general health check up in a Japanese population[62]. The rate of abnormal RSI scores (RSI>13) was seven percent, with the highest rate found in middle aged males with the highest BMI (26 – 28, which is perhaps not high compared to a UK or USA population).

The objective for Chapter 3 is to assess the relationship between the different PROMs and clinician rated throat appearances with one another.

3.2 Methods

The relationship analysis included age, BMI, RSI, CReSS, LPR-HRQL (overall impact of acid reflux score) and RFS. To assess whether the questionnaires' data were related, the raw data was plotted graphically as scatter diagrams. The relationship between the questionnaires and the RFS was assessed graphically using (bivariate) pairwise scatter plots. The relationship was reported using the Pearson correlation coefficient, which is a measure of the strength of the linear correlation between two variables. The strength of correlation has been described by Evans as: 0.00-0.19 "very weak", 0.20-0.39 "weak", 0.40-0.59 "moderate", 0.60-0.79 "strong", 0.80-1.0 "very strong"[63]. Further descriptive statistics and graphical representations were considered to aid in the definition of any relationships that emerged.

3.3 Results

Table 3-1 displays the correlation results of age, BMI, RSI, CReSS and LPR-HRQL with one another. Increasing age was significantly negatively correlated with both the CReSS and LPR-HRQL, but not with the RSI. Patients' BMI was not related to age or any of the three PROMs. The RSI was strongly positively correlated with the CReSS and moderately positively

correlated with the LPR-HRQL. The CReSS and LPR-HRQL were strongly positively correlated. These relationships are displayed graphically in Figure 3-1. Table 3-2 shows the correlation of the RFS with the age, BMI and the three questionnaires. The RFS was weakly positively correlated with BMI. There was no relationship between the RFS and any of the 3 symptoms questionnaires

Table 3-1 Pairwise correlations between of questionnaire outcomes, Age and BMI (n=328)

Variables	Age	BMI	RSI	CReSS	LPR-HRQL
Age	1				
BMI	0.08	1			
RSI	-0.08	-0.03	1		
CReSS	-0.24**	0.02	0.73**	1	
LPR-HRQL	-0.29**	-0.03	0.58**	0.71**	1

** Correlation is significant at the 0.01 level (2-tailed).

Figure 3-1 scatterplot depicting correlation between questionnaire outcomes and Age / BMI

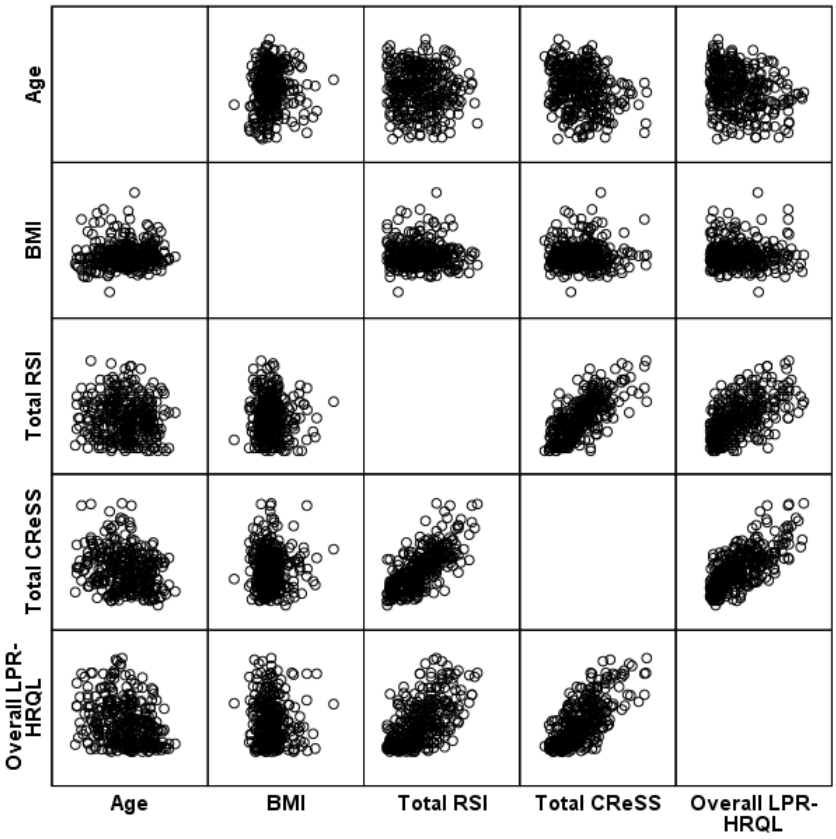


Table 3-2 Correlation of RFS with Age, BMI and the questionnaires (n=245)

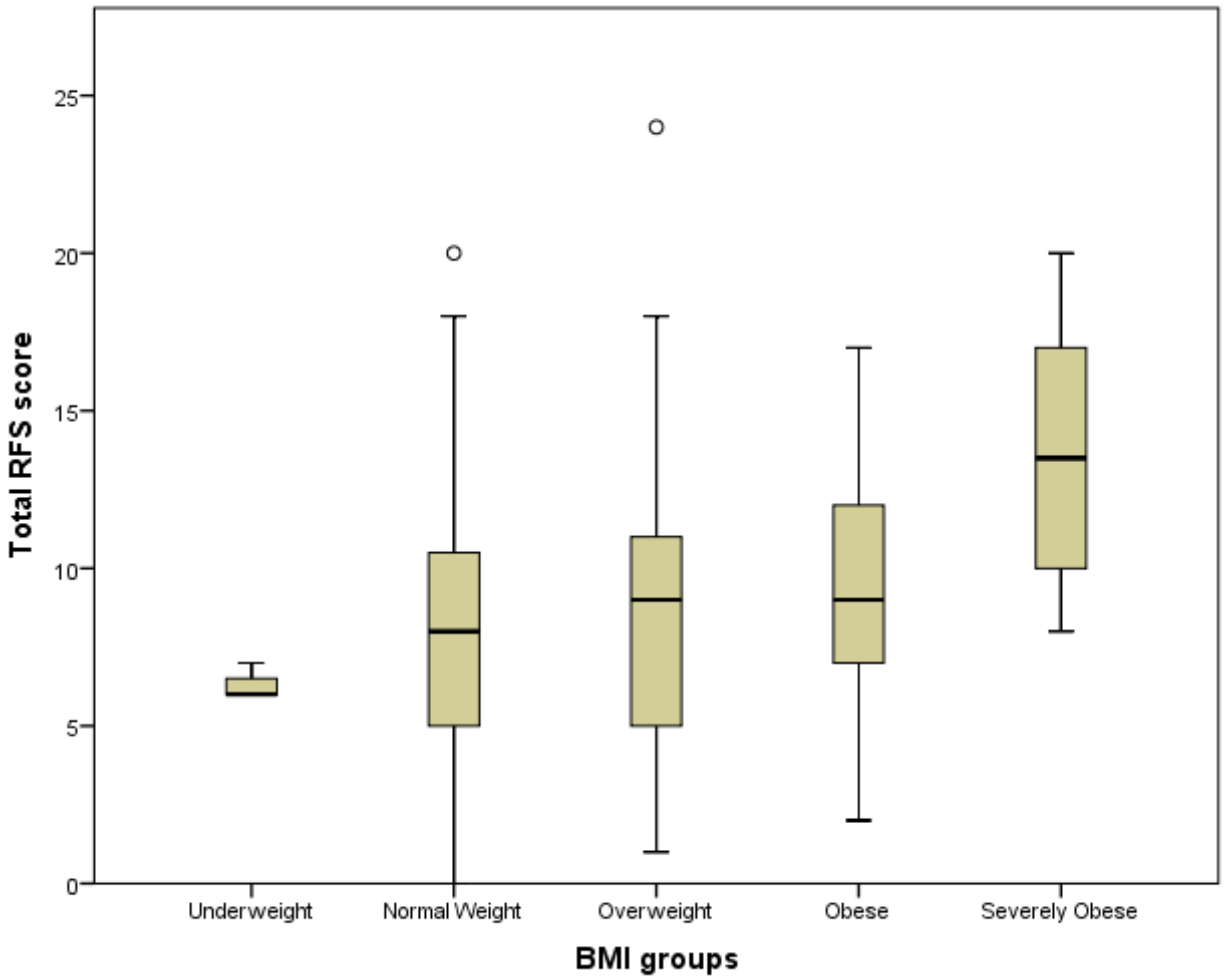
	Age	BMI	RSI	CReSS	LPR-HRQL
RFS	0.15*	0.25**	0.06	0.04	-0.02

* Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

The relationship between the total RFS score and BMI was explored further with descriptive analyses and graphically with a box plot (Figure 3-2). Patients BMI were categorised into "underweight" (BMI <19), "normal weight" (19 – 24.9), "overweight" (25 – 29.9), "obese" (30 – 39.9), "severely obese" (> 40)[64]. Of 325 patients with BMI data, three were underweight (0.9%), 91 (27.4%) were normal weight, 133 (40.1%) were overweight, 83 (25%) were obese, and 15 (4.5%) were severely obese.

Figure 3-2 Box plot of BMI categories versus Total RFS score



3.4 Discussion

The TOPPITS baseline data is comparable to the other studies in suggesting lower symptom scores with increasing age. This observation has been demonstrated previously with the RSI [27, 60]. However, when analysing the TOPPITs data, increasing age was related to a lower score with the CReSS and LPR-HRQL but not the RSI. These findings should not be viewed as implying that throat symptoms are less troublesome as patients get older. The questionnaires used are subjective measures of symptom burden; how individual patients score a symptom on a Likert scale is determined by themselves alone. One plausible explanation for these results could be that as patients become older, they develop co-morbidities that may alter their perception of their throat symptoms relative to other symptoms. It is more likely that as patients grow older, the scores they report are lower, rather than them reporting fewer symptoms with increasing age. BMI was not related to any of the three symptom questionnaires but there was a suggestion of a weakly positive correlation with the RFS (discussed further below).

As may have been expected, the three symptom questionnaires were related in scoring outcomes. The RSI and CReSS were strongly correlated, likewise the CReSS and LPR-HRQL, and the RSI and LPR HRQL showed moderate correlation. Whilst there are specific criteria, as set out in section 1.3.1 (Instrument validation in throat symptoms), to aid validation of instruments, the strong correlation between the RSI and CReSS probably suggests that only one of these questionnaires need be utilised. The same could be argued for the LPR-HRQL; perhaps only one questionnaire is required.

No linear relationship was observed between the RFS and the three questionnaires in this study. In comparison to the study by Mesallam et al [26] with 40 patients, the TOPPITS baseline data contained RFS and symptoms outcomes for 245 patients. Only one blinded assessor of the images was used in TOPPITS; an experienced speech and language professor. This ensured a consistent approach to the scoring across a large sample. The RFS could be considered a subjective measure. If the single assessor's approach was to score relatively low this could have affected the results. Given the importance of this finding, it may be appropriate to repeat the scoring with further blinded assessors and reassess whether the results are consistent. However, the lack of any relationship between the RFS and questionnaire scores from the TOPPITS population is very clear. As discussed in the thesis introduction, the Mesallam et al[26] study had some potential methodological flaws and was undertaken on a much smaller group of patients. Unfortunately, the equipment used to capture the throat picture was not standardised across the recruiting centres. However, in stark contrast to Mesallam et al no relationship was found in the TOPPITS baseline data between symptoms and throat / voice box appearances. These findings raise the question of whether throat appearances should be used in the assessment and management of patients with persistent throat symptoms. Whilst a weak relationship was noted between the RFS and BMI, on further analysis it appears that the relationship is influenced by the small "severely obese" group of patients, comprising only 15 patients. The box plot usefully demonstrates that the total RFS scores are similar across the larger BMI category groups. This relationship is unlikely to be of clinical value.

3.5 Conclusion

This analysis of the TOPPITS baseline data has shown that the three questionnaires are related in terms of total scores and that increasing age is associated with reduced symptom reporting. There was no relationship observed in this dataset between the RFS throat scores and any of the three symptoms questionnaires.

Chapter 4 Exploratory Factor Analysis

4.1 Introduction

Thus far the thesis has described the demographics of the TOPPITS population and descriptive analyses of the outcomes data. It has demonstrated that the scoring patterns over the three symptom questionnaires are positively related and that the symptom scores do not correlate with throat appearance. This chapter will explore whether the large number of symptoms, for which data exists, can be reduced into more simplified categories that may be of clinical utility. If dimension reduction into fewer defined categories is appropriate, these categories will then be used in Chapter 5 to explore whether patients can be placed into distinct groups as defined by their presenting symptoms.

Exploratory factor analysis (EFA) is a statistical method used to identify latent constructs or factors within a set of measured variables, which groups correlated variables together. Amongst a set of measured variables there may be unmeasured reasons as to why certain variables correlate or covary; latent constructs. The methodology can reduce a large set of items to a reduced number of factors, with each factor comprised of a number of correlated items. EFA has no *a priori,* or predefined, theory as to which items are expected to map to which factors.

The principal aim of this thesis is to explore whether a more clinically meaningful classification of patient symptoms can be proposed. TOPPITS has provided a wealth of patient reported symptoms that can be analysed as variables. Any clinically meaningful classification of symptoms would ideally have a small number of groups within it. Using EFA may allow an appropriate reduction of the unwieldy number of symptoms into fewer manageable hypothetical groups, which could then be used as a means to explore whether distinct groups of patients can be defined with cluster analysis.

EFA methodology has been used to analyse the three questionnaires, in separate analyses, in previously published studies. These are described in further detail within the discussion section. Cathcart et al [50] and Printza et al[49] identified two factors within the RSI, both studies demonstrating similar individual items grouped together within these two factors. In the former study the two factors comprised:

1. Breathing difficulties or choking, coughing after eating or lying down, troublesome cough, difficulty in swallowing, heartburn or indigestion.

2. Throat clearing, globus, postnasal drip, and hoarseness or voice disorder.

Drinnan et al [45] found three factors within the CReSS questionnaire which were subsequently defined as: a “gastrointestinal (oesophageal)” factor; an “upper airway factor” – relating to cough, breathing, mucus and hoarseness; and an “obstruction / choking and globus” factor (pharyngeal or throat). The original study which described the LPR-HRQL questionnaire found that each of the five domains had a single factor structure, other than the voice domain which was affected by a single reversed scoring item[57], implying that the questions in each domain were appropriately grouped. Confirmatory factor analysis is another method of factor analysis which assesses how well a study population fits within a model of predefined factors. Confirmatory factor analysis was considered for this thesis. However, given the relative paucity of EFA studies in persistent throat symptom populations and the fact this thesis had an opportunity to analyse three outcome measures combined, EFA was chosen to explore the potential factors rather than seeking to confirm previously published factors. EFA was chosen over other data-reduction methods such as Principal Component Analysis (PCA). While PCA is useful for reducing the number of variables (but does not give insight into latent factors), EFA identifies and measures variables that cannot be measured directly (i.e. latent variables – the clinically meaningful underlying symptoms). Using EFA also allows comparison with the published literature outlined above.

The objective of this chapter was to explore dimension reduction with the aim of identifying a smaller number of meaningful clinical factors. This will be achieved using EFA, firstly by conducting an EFA on combined data from the three questionnaires, and secondly by conducting an EFA of the three questionnaires individually, in order to allow comparison with the previously published studies.

4.2 Methods

The following methods are built on recommendations from Costello and Osborne[65]. A series of web based tutorials written and filmed by James Gaskin [66-69] proved to be a very accessible resource for the practical implementation of these methods.

Exploratory Factor Analysis

Data from the RSI, CReSS and LPR-HRQL were combined for the EFA. The data was cleaned throughout the TOPPITS trial by a team of data managers who raised queries with sites and chased up missing data wherever possible. All nine items and all 34 items from the RSI and CReSS were included. Only the voice, cough, throat and swallow domain items (29

in total) were included from the LPR-HRQL. Whilst the "overall impact of acid" items were included in the original construct validity work[57], it was felt that these items were too varied in their nature to reasonably comprise one set of related questions. The overall thermometer questions that follow each domain were similarly not included. Patients with missing data in any one of the items included from the three questionnaires were excluded from the analysis.

EFA was conducted in SPSS version 24. The commands used are detailed in Appendices - EFA Methodology in SPSS. The key elements of an EFA are:

1. Assessing the data to confirm that data reduction methods are appropriate.

2. Choosing the correct method to extract the factors.

3. Deciding how many factors to retain for the final model. The use of Eigenvalues and Scree Plots are widely used techniques. EFA involves calculations using mathematical matrices. Eigenvalues and Eigenvectors are components used to transform mathematical matrices of data values.

4. Choosing the correct rotation method. If all data items are plotted on an x and y graph, rotating the axes can allow the data to be more easily interpreted. Two rotation methods exist: orthogonal which rotates the axes around the intersection of x and y, keeping x and y at right-angles; and oblique rotation which allows the angles between x and y to vary from 90 degrees. Orthogonal is more appropriate for uncorrelated data and oblique for correlated data.

5. Assessing the strength of association of each variable within the factors, and deciding on loading values above which items are kept within the factor structure. Factor loadings describe the correlation between each variable and the underlying factor – the higher the value between zero and one, the stronger the correlation.

In the first step, variables in a factor analysis should be linearly correlated with each other and Bartlett's test of sphericity was used to check for the presence of correlation between variables. The Kaiser-Meyer-Olkin (KMO) test is used to measure the sampling adequacy in data. Large KMO values (on a scale of zero to one) are desirable as this implies that correlations between pairs of variables (potential factors) can be explained by other variables.

In step two, the Principal Axis Factoring extraction method was used, as it is appropriate for non-normal data. A non-normal extraction method seemed appropriate considering that Chapter 2 showed that the individual items of the RSI, when plotted as a histogram with overlying normal curve (see Appendices - Demographic Data Analysis), did not appear graphically as normally distributed. In addition, the construction of a Likert scoring scale for items on the questionnaires means that, for example, a rise from one to two may not necessarily be equivalent to a rise from a score of three to four.

In step three, factors were retained with an eigenvalue greater than one. Additionally, scree tests were used to plot the factors' eigenvalues graphically. Factors lying above the "break" or "shoulder" of the curve were retained. Where there was ambiguity if a factor was at or near the break in the curve, several factor analyses were run with a differing number of factors to assess which produced the "cleanest" pattern matrix; with individual questionnaire items loading on to a single factor.

In step four, Oblique Promax Rotation of the data was used, given the large dataset and the expected relationship of the factors, supported by the strong correlation demonstrated between the questionnaires in Chapter 3.

In step five, the suppression of small coefficients was explored where appropriate, to remove items with low loading values. Items' communalities scores were displayed as a measure of correlation with other items. Individual items should be > 0.2-0.4 with low to moderate communalities acceptable in the 0.4 to 0.7 range. The rotation of the model was used to assess if it had converged within 10 iterations, denoting an acceptable EFA. The total variance of the model was used to explain how much of the total variance the extracted factors explain.

The individual items loading onto the factors were displayed within pattern matrices, with the loading value of each item per factor displayed. The adequacy of the EFA models was considered by assessing for results with no low loading scores (defined in this instance as <0.4). If a single item loaded on more than one factor, this was deemed a cross loading item. This was considered acceptable if the loadings differ by more than 0.2 across factors. Cross loading items were removed and the EFA re-run, if the factor had several strong loading items (>0.5). A factor with less than three items was considered to be weak, with five or more strongly loading items being ideal. If a clean factor structure was produced, the individual factors were named – either according to the specified questions in the tool, or as new latent variable groups. The saved factor variables were scored using the regression method, given the oblique rotation used. The saved factor score is a standardised score, with a mean for all

patients of zero. This factor score can then be utilised as a predictor variable for future analyses.

4.2.1 *Sample size*

Exploratory factor analysis solutions can be problematic with small sample sizes that contain items with low communalities scores or low loading values within factors. A general rule of thumb that is commonly applied is a ratio of 10 subjects to 1 item[65]. However, Mundfrom et al[70] cite a range of recommended ratios from 3:1 subjects: variables to 10:1, and some suggestions including a minimum overall sample size - which varies widely. They show that the necessary sample size reduces as the ratio of variables to factors increases – i.e. more variables. Even with low communalities, if the ratio of variables to factors is 7 or more the sample size is no more than 180 individuals. The review does not offer a conclusion on the most appropriate sample size to use. Costello and Osborne[65] state that "Strong data in factor analysis means uniformly high communalities without cross loadings, plus several variables loading strongly on each factor" and to achieve this requires that EFA is a large sample technique. The sample size for TOPPITS was calculated based on the primary RSI outcome measure at 16 weeks. In this respect, the sample size available for the EFA was pre-defined by the trial. However the sample size of TOPPITS (328 patients with no missing data for all three questionnaires) offers the opportunity to improve the evidence from EFA of the 3 questionnaires as it is far larger than the other EFA studies previously cited.

4.2.2 *Reliability Analysis*

The items in each factor were analysed using Cronbach's alpha test and the mean factor loading score calculated. Cronbach's alpha gives a measure of how closely related a set of variables are. Reliability analysis was performed and assessment made on the factor strength if items were removed from the factor. The mean factor loading score was calculated for each factor separately, with a value > 0.7 recommended to represent an appropriate factor structure[68].

4.2.3 *Separate Questionnaire EFAs*

Separate EFAs for the three individual questionnaires were conducted following the methodology described above. This will give the opportunity to compare the TOPPITS population EFA with the previously published results.

4.2.4 *Abbreviations used in EFA tables*

The RSI and CReSS symptoms are included within the results table. To aid editing, the longer LPR-HRQL items are abbreviated within the results table (Table 4-1).

Table 4-1 LPR-HRQL abbreviations

lprVOICE 1	My Voice Problems Make It Difficult For Me To Work
lprVOICE 2	I Feel Satisfied With The Way My Voice Sounds (Reverse scoring)
lprVOICE3	Being Hoarse Makes It Hard For Me To Communicate My True Self
lprVOICE4	My Voice Makes Others Feel Uncomfortable To Listen To Me
lprVOICE5	I Can't Sing As Much As I Would Like To Because Of My Voice
lprVOICE6	I Find It Hard To Meet New People Because Of What They Will Think
lprVOICE7	The Sound Of My Voice Makes People Think I'm Angry Or Upset
lprVOICE8	Straining To Talk Is Tiring
lprVOICE9	I Feel Embarrassed About The Sound Of My Voice
lprVOICE10	I Avoid Talking Because Of The Effort
lprVOICE11	My Voice Problems Make It Difficult For Me To Do My Job
lprVOICE12	I Am Afraid I Might Lose My Voice Forever
lprCOUGH14	My Coughing Embarrasses Me
lprCOUGH15	I Avoid Social events Where My Coughing Might Bother Others
lprCOUGH16	I Have To Leave The Room Because Of My Coughing
lprCOUGH17	People Think I Am Sick Because Of My Coughing
lprCOUGH18	My Co-workers Can Hear Me Coming Because Of My Coughing
lprCOUGH19	I Worry About Having A Coughing Spell At A Bad Time
lprTHROAT21	People Notice How Much I Have To Clear My Throat
lprTHROAT22	Clearing My Throat Has A Negative Effect On Sex
lprTHROAT23	Clearing My Throat Has A Negative Effect On Friendships
lprTHROAT24	The Need To Clear My Throat Makes It Hard To Talk
lprTHROAT25	I Feel Frustrated About Having To Clear My Throat So Often
lprTHROAT26	I Avoid Social Events Because Of The Need To Clear My Throat
lprSWALLOW28	I Feel A Lump In My Throat Which Makes It Hard To Swallow
lprSWALLOW29	I Hesitate To Eat In Public Because I Have Trouble Swallowing
lprSWALLOW30	I Am Afraid Of Choking In My Sleep
lprSWALLOW31	I Am Bothered By A Burning Feeling In My Throat
lprSWALLOW32	I Awaken From Sleep Gasping For Breath

4.3 Results

4.3.1 *EFA with factors defined by Eigenvalues > 1*

Twelve patients were excluded from the analysis due to missing data. The EFA was performed on data from 332 patients. The combined questionnaires EFA had a subject to variable ratio of 332:72. The second item in the voice domain of the LPR-HRQL (LPR VOICE 2) had a very low communality value, at 0.20. This may have been a reflection of its reverse scoring – the only reverse scoring item in the questionnaires. Therefore this item was removed from the analysis.

The KMO measure of sampling adequacy was 0.91 and Bartlett's test of sphericity suggests a statistically significant linear relationship between variables (approximate Chi-square = 15832, $p<0.001$) and thus a factor analysis may provide useful insights into this data. The majority of included items had satisfactory communality scores (Table 4-2). There were borderline scores for the CReSS items: 'hiccups', 'vomiting', 'bad breath', 'headache', 'back pain', and for the fifth LPR-HRQL voice item ('I can't sing as much as I would like to because of my voice'). These items were maintained in the initial EFA.

Table 4-2 Communalities Scores for EFA Combined Questionnaires - Unspecified number of Factors

Questionnaire Item	Communality Extraction
RSI Hoarseness or problem with voice	0.73
RSI Throat Clear	0.56
RSI Excess Throat Mucus or PND	0.66
RSI Difficulty Swallowing Food Liquids or Tablets	0.74
RSI Cough after Eating or Lying	0.72
RSI Breathing Difficulties or Choking Episodes	0.62
RSI Troublesome Cough	0.61
RSI Something Caught or Lump In Throat	0.63
RSI Heartburn Chest Pain Indigestion or Stomach Acid Coming Up	0.73
CReSS Heartburn	0.75
CReSS Pressure in Chest	0.47
CReSS Regurgitation	0.50
CReSS Acid Sour Mouth	0.58
CReSS Gurgling Stomach	0.48
CReSS Lump in Throat	0.69
CReSS Difficulty Swallowing Food	0.82
CReSS Difficulty Swallowing Liquids	0.55
CReSS Nausea	0.49
CReSS Pain in Throat	0.46
CReSS Vomiting	0.34
CReSS Bloating	0.47
CReSS Belching	0.61
CReSS Flatulence	0.56
CReSS Hiccups	0.37
CReSS Decreased Appetite	0.53
CReSS Rush of Saliva	0.45
CReSS Feeling Full Early	0.57
CReSS Bad Breath	0.32
CReSS Back Pain	0.34
CReSS Headache	0.29
CReSS Choking	0.55
CReSS Cough Upright	0.73
CReSS Cough after Eating	0.62
CReSS Cough Lying	0.62
CReSS Wheezing	0.55
CReSS Difficulty Breathing	0.74
CReSS Hoarseness	0.78

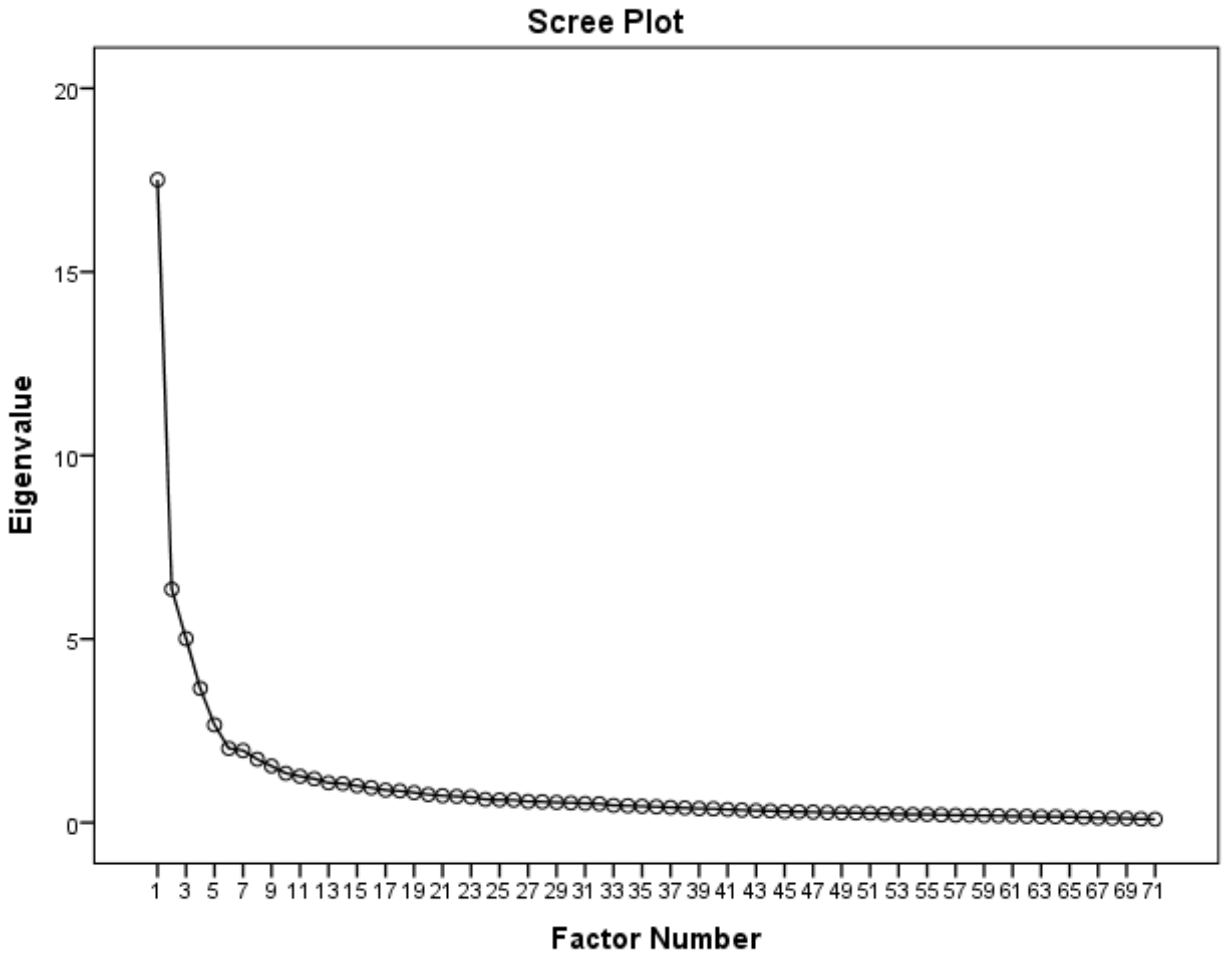
CReSS Throat Clearing	0.69
CReSS Excess Mucus	0.78
CReSS Mucus Dripping	0.72
CReSS Feeling Things Stuck Throat	0.64
CReSS Indigestion	0.70
CReSS Stomach Acid Up	0.77
lprVOICE 1	0.70
lprVOICE 3	0.73
lprVOICE 4	0.71
lprVOICE 5	0.40
lprVOICE 6	0.68
lprVOICE 7	0.60
lprVOICE 8	0.74
lprVOICE 9	0.77
lprVOICE 10	0.73
lprVOICE 11	0.70
lprVOICE 12	0.56
lprCOUGH 14	0.64
lprCOUGH 15	0.70
lprCOUGH 16	0.68
lprCOUGH 17	0.74
lprCOUGH 18	0.64
lprCOUGH 19	0.75
lprTHROAT 21	0.60
lprTHROAT 22	0.68
lprTHROAT 23	0.65
lprTHROAT 24	0.59
lprTHROAT 25	0.67
lprTHROAT 26	0.58
lprSWALLOW 28	0.60
lprSWALLOW 29	0.59
lprSWALLOW 30	0.55
lprSWALLOW 31	0.64
lprSWALLOW 32	0.58

The initial EFA model explained a total variance of 61.6%, and comprised 15 factors. It was not a clean factor structure. The scree plot suggested an optimal number of factors between 6 and 10 (Figure 4-1). The EFA model was incrementally improved through a series of 15 steps, removing problematic loading items at each stage, with each change resulting in a

cleaner factor model. The individual steps and the items removed are detailed in "EFA of three questionnaires with factors defined by Eigenvalues > 1" of the Appendix.

This then gave a 10 factor model (see appendix Table 7-1) with several items cross loading onto more than one factor. However, all the cross loading items had a strong loading score on one factor and low scores on the cross loading factor. These items were preserved therefore in the model.

Figure 4-1 Scree plot of items' eigenvalues – Combined questionnaires EFA with unspecified number of factors



4.3.2 *Defined number of variables EFA*

The above methodology was repeated up to the resultant 10 factor model stage. The number of factors was defined as seven rather than being based on eigenvalues greater than one. The scree plot (Figure 4-2) does not have a clear shoulder, but would suggest between six and eight factors may be a reasonable number to include. The extraction method remained as

Principal Axis Factoring, with Promax rotation with Kaiser Normalisation. Small coefficients were supressed to > 0.4 values.

This produced a model with a KMO of 0.90. All communalities were greater than 0.44 (see appendix Table 7-2). The total variance explained by this model was 55%. The rotation converged in 9 iterations. 'CReSS Wheezing' and 'CReSS flatulence' did not load with scores >0.40 and were removed.

Figure 4-2 Scree plot for combined questionnaires EFA - seven factor model

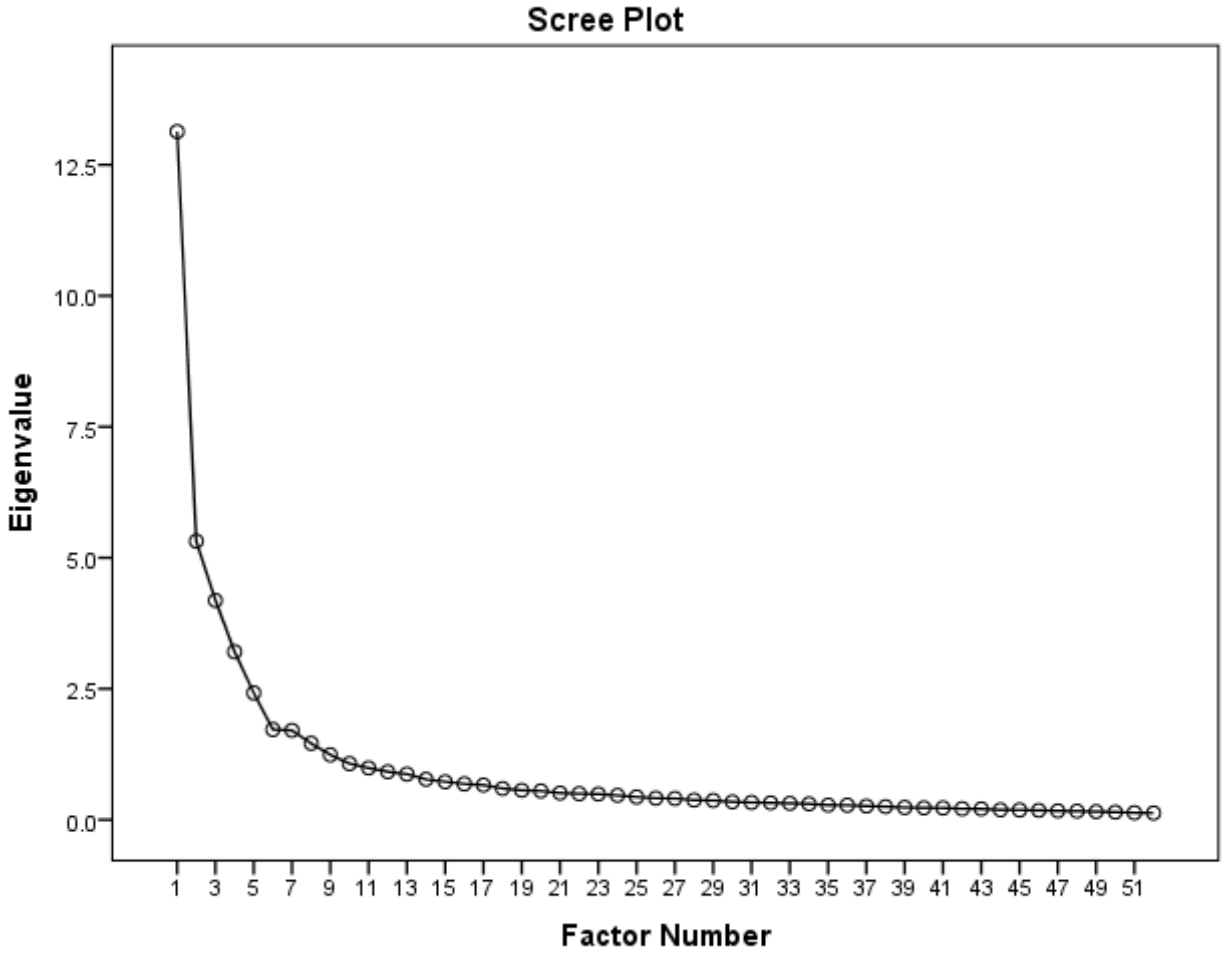


Table 4-3 Pattern Matrix for seven Factor model

Questionnaire Item	Factor						
	1	2	3	4	5	6	7
lprVOICE 8	0.90						
lprVOICE 3	0.88						
lprVOICE 9	0.85						
lprVOICE 4	0.81						
lprVOICE 10	0.78						
CReSS Hoarseness	0.76						
lprVOICE 1	0.75						
RSI Hoarseness	0.75						
lprVOICE 6	0.68						
lprVOICE 12	0.65						
lprVOICE 5	0.60						
lprVOICE 7	0.60						
RSI Troublesome Cough		0.85					
CReSS Cough Upright		0.85					
RSI Cough after Eating or Lying		0.83					
CReSS Cough Lying		0.78					
CReSS Cough after Eating		0.71					

Questionnaire Item	Factor						
	1	2	3	4	5	6	7
lprCOUGH 17		0.69					
lprCOUGH 19		0.67					
lprCOUGH 16		0.62					
lprCOUGH 18		0.53					
CReSS Indigestion			0.87				
RSI Heartburn,ChestPain, Indigestion,Stomach Acid			0.87				
CReSS Stomach Acid Coming Up			0.86				
CReSS Heartburn			0.86				
CReSS Acid Sour Mouth			0.63				
CReSS Regurgitation			0.54				
CReSS Belching			0.53				
CReSS Bloating			0.46				
CReSS Excess Mucus				0.77			
CReSS Throat Clearing				0.74			
RSI Excess Throat Mucus or PND				0.71			
RSI Throat Clear				0.70			
CReSS Mucus Dripping				0.67			
lprTHROAT 25				0.56			
lprTHROAT 21				0.50			
CReSS Difficulty Swallowing Food					0.70		
lprSWALLOW 32					0.64		

Questionnaire Item	Factor						
	1	2	3	4	5	6	7
CReSS Choking					0.60		
lprSWALLOW 30					0.58		
RSI_Difficulty Swallowing Food Liquids or Tablets					0.57		
CReSS Difficulty Swallowing Liquids					0.57		
CReSS Difficulty Breathing					0.53		
RSI Breathing Difficulties or Choking Episodes					0.46		
RSI Something Caught or Lump In Throat						0.68	
CReSS Feeling Things Stuck Throat						0.62	
CReSS Lump in Throat						0.62	
lprTHROAT 23							0.70
lprTHROAT 22							0.58
lprTHROAT 26							0.53

4.3.3 *Re-labelling of Factors*

Examining the question items within each factor reveals that the symptoms do seem to link together logically. It is useful to give the factors more clinically meaningful names.

Factor 1	Voice Factor
Factor 2	Cough Factor
Factor 3	Gastrointestinal Symptoms (GI) Factor
Factor 4	Airway Symptoms and Dysphagia Factor
Factor 5	Throat Clearing Factor
Factor 6	Lump in Throat Factor
Factor 7	Life Events Factor

4.3.4 *Reliability Analysis*

Table 4-4 shows the reliability analysis for the seven factor EFA model. The correlation of variables within each factor and the mean loading score appeared stronger the greater the number of variables contained in each factor. Only one factor would improve, in terms of correlation of variables, if items were removed. The seventh factor of Life events would increase from Cronbach's alpha of 0.77 to 0.80 if LPR Throat 26 was removed (I avoid social events). This improvement in the Cronbach's alphas score does not seem significant enough to justify dropping the item, which would leave a factor with only two items. The fact that the removal of any other items would not improve the correlation of variables with the factors suggests the EFA model is appropriate.

Table 4-4 Reliability analysis for the seven factor EFA model

Factor number	Factor label	Cronbach's alpha score	Mean factor loading score	Factor structure improved with item(s) removed?
1	Voice	0.94	0.75	No
2	Cough	0.92	0.73	No
3	Gastrointestinal	0.89	0.70	No
4	Airway / Dysphagia	0.83	0.59	No
5	Throat clearing / Mucus	0.85	0.65	No
6	Life events	077	0.63	Yes
7	Lump	0.81	0.61	No

4.3.5 *Reflux Symptom Index Exploratory Factor Analysis*

An EFA was conducted for the RSI using the same methods as described above, using Principal Axis Factoring with small coefficients suppressed to >0.2, with the number of factors defined by Eigenvalues >1.

The KMO was 0.61 with 41% of the total variance explained (i.e. appreciably lower than the values for the combined EFA approach). Communalities were < 0.2 for 3 items (Table 4-5). The scree plot did not have shoulder to the curve to indicate a clear number of factors to include (Figure 4-3). An initial 3-factor model was produced (Table 4-6). This was not a clean factor model, with cross loading scores and low scores evident. The cough items had high loading scores in factor 1, difficulty swallowing in factor 2 and excess mucus in factor 3. The remaining 5 items in the RSI had low loading scores.

Table 4-5 Communalities for RSI EFA three factor model

RSI Item	Extraction Communality
RSI Hoarseness or problem with voice	0.12
RSI Throat Clear	0.21
RSI Excess Throat Mucus or PND	0.76
RSI Difficulty Swallowing Food Liquids or Tablets	0.56
RSI Cough after Eating or Lying	0.60
RSI Breathing Difficulties or Choking Episodes	0.39
RSI Troublesome Cough	0.81
RSI Something Caught or Lump In Throat	0.15
RSI Heartburn Chest Pain Indigestion or Stomach Acid	0.10

Figure 4-3 Scree Slope for RSI EFA with three factors

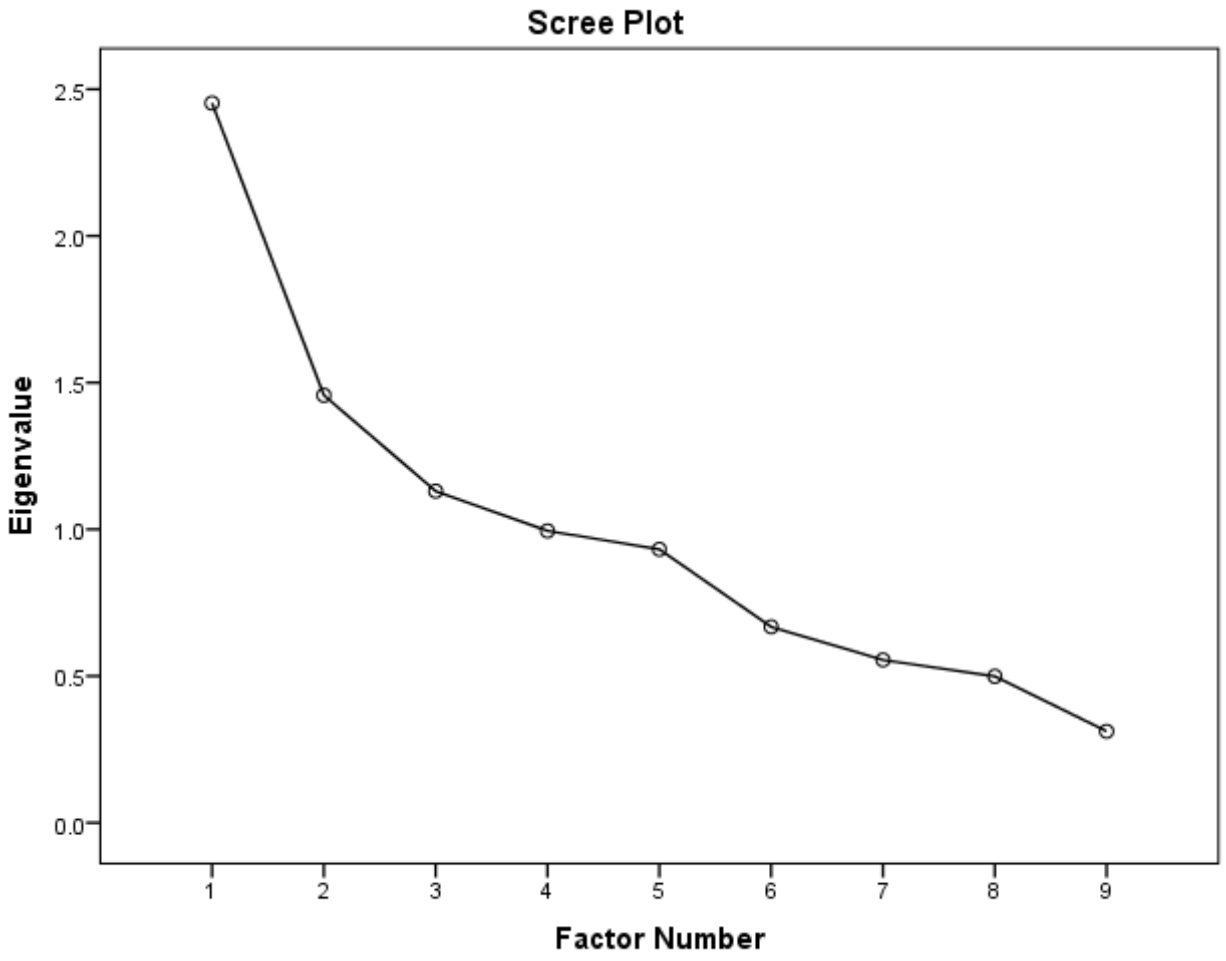


Table 4-6 RSI EFA three Factor Pattern Matrix

RSI items	Factor		
	1	2	3
RSI Troublesome Cough	0.93	-0.21	
RSI Cough after Eating or Lying	0.66	0.25	
RSI Difficulty Swallowing Food Liquids or Tablets		0.77	
RSI Breathing Difficulties or Choking Episodes	0.28	0.46	
RSI Something Caught or Lump In Throat		0.38	
RSI Excess Throat Mucus or PND			0.90
RSI Throat Clear	0.21		0.35
RSI Heartburn Chest Pain Indigestion or Stomach Acid			0.25
RSI Hoarseness or problem with voice			0.22

With the number of factors defined as two, the KMO remained at 0.61, with low communalities as above for three items. Only 31% of the total variance was explained by this model. The scree plot lacked a clear shoulder. The pattern matrix (Table 4-7) did not show a clean structure, with only the cough items in factor 1 having high loading scores, and difficulty swallowing in factor two. The other six items had low loading score, with two cross loading and the heartburn item failing to load above 0.2.

Table 4-7 RSI two factor EFA Pattern Matrix

	Factor	
	1	2
RSI_TroublesomeCough	.78	
RSI_Cough_afterEating_or_Lying	.68	.21
RSI_ThroatClear	.42	
RSI_ExcessThroatMucus_orPND	.37	
RSI_Hoarseness_or_problem_with_voice	.22	
SI_DifficultySwallowingFood_Liquids_orTablets		.81
RSI_BreathingDifficulties_orChokingEpisodes	.36	.43
RSI_SomethingCaught_orLumpInThroat		.36
RSI_Heartburn_ChestPain_Indigestion_orStomachAcidComingUp		

4.3.6 *CReSS EFA*

Using the defined EFA methodology, the KMO was 0.89 with communalities for all items > 0.3 (Table 7-3), and total variance explained of 52%. Table 7-4 shows the initial pattern matrix with factors extracted with Eigenvalues >1.

The scree plot is shown in Figure 7-10 and lacks a clear shoulder to define the number of factors to include. With the number of factors suppressed to three, in line with Drinnan et al[45], the model gave a KMO of 0.89, with similarly high communalities as above. The total variance explained was 37%. The pattern matrix is shown in the appendix Table 7-5.

Cross loading factors with low scores were removed or suppressed by raising the small coefficient score to 0.3, to give a cleaner three factor model (Table 4-7). This removed: ‘feeling full early’, ‘pressure in chest’, ‘vomiting’ and the low scoring cross loadings were suppressed: ‘back pain’, ‘hiccups’ and ‘bad breath’. The KMO for this model was 0.89 with the total variance explained of 38%. The ‘back pain’ and ‘hoarseness’ item communalities were slightly below 0.3, but remained satisfactory. The resulting three factor model could be re-labelled as factor one – Gastroesophageal symptom, factor two – Upper Airway symptoms, and factor three – Pharyngeal symptoms.

Table 4-8 CReSS three Factor Cleaned EFA Pattern Matrix

Questionnaire Item	Factor		
	1	2	3
CReSS_Indigestion	0.90		
CReSS_Stomach acid up	0.87		
CReSS_Heartburn	0.82		
CReSS_Acid or sour taste in mouth	0.67		
CReSS_Regurgitation	0.61		
CReSS_Belching	0.59		
CReSS_Bloating	0.55		
CReSS_Gurgling stomach	0.51		
CReSS_Flatulence	0.40		
CReSS_Nausea	0.38		
CReSS_Rush of saliva in mouth	0.34		
CReSS_Coughing when upright		0.86	
CReSS_Cough when lying		0.75	
CReSS_Coughing after eating		0.72	
CReSS_Wheezing		0.65	
CReSS_Mucus dripping in throat		0.51	
CReSS_Excess mucus		0.51	
CReSS_Difficulty breathing		0.47	
CReSS_Throat clearing		0.44	
CReSS_Hoarseness		0.34	
CReSS_Difficulty swallowing food			0.86
CReSS_Difficulty swallowing liquids			0.74
CReSS_Lump in throat			0.59
CReSS_Feeling things stuck throat			0.55
CReSS_Pain in throat			0.47
CReSS_Choking		0.34	0.46
CReSS_Decreased appetite			0.45
CReSS_Headache			0.33

Given the reproducible results from the CReSS questionnaire EFA when compared to the published literature, a three factor model for all three questionnaires was performed. The EFA methodology was repeated with suppression of coefficients to 0.3. Whilst three factors could be demonstrated, the CReSS items were distributed differently to the EFA of this questionnaire alone. The 3 factors were Cough and Throat Clear factor, a Voice factor, and a Gastrointestinal factor. The CReSS hoarse item was separate to the other CReSS items which are distributed among the other two factors. This three factor model for all three questionnaires did not appear as applicable as the 7 factor model detailed above.

4.3.7 *LPR-HRQL EFA*

Following the above EFA methodology, the initial model based on factor extraction with eigenvalues greater than one gave a KMO of 0.91, communalities were all > 0.3, and the total variance explained was 59%. Table 4-8 shows the initial pattern matrix.

'Throat 24' cross loaded with the voice items ('the need to clear my throat makes it hard to talk'). Removing this item did not alter the pattern matrix significantly. The Throat items remained split over two factors.

Table 4-9 LPR-HRQL EFA Pattern Matrix					
Questionnaire Item	**Factor**				
	1	**2**	**3**	**4**	**5**
lprVOICE8	0.91				
lprVOICE3	0.86				
lprVOICE9	0.84				
lprVOICE1	0.83				-0.38
lprVOICE10	0.81				
lprVOICE11	0.80				-0.25
lprVOICE4	0.78				
lprVOICE6	0.72				0.29
lprVOICE12	0.66				
lprVOICE5	0.61				
lprVOICE7	0.58				0.21
lprCOUGH19		0.87			
lprCOUGH16		0.87			
lprCOUGH17		0.85			
lprCOUGH18		0.71			
lprCOUGH14		0.66		0.24	
lprCOUGH15		0.64			0.32
lprSWALLOW30			0.73		
lprSWALLOW29			0.63		
lprSWALLOW28			0.60		
lprSWALLOW32			0.58		
lprSWALLOW31			0.55		
lprTHROAT25				0.81	
lprTHROAT21				0.70	
lprTHROAT24	0.38			0.41	
lprTHROAT23					0.67
lprTHROAT22					0.62
lprTHROAT26					0.54

4.4 Discussion

This chapter significantly adds to the evidence base on latent constructs within throat symptoms in questionnaire reporting. By combining the RSI, CReSS and LPR-HRQL questionnaires together within an EFA, it offers the opportunity to compare the factor structure, over many symptom items within a large sample size, with the factor structure observed within the individual questionnaires. However, there is great overlap between the questionnaires in terms of symptoms, and the relationship analysis performed in Chapter 3 showed the strong correlation in scoring between the tools. The subject to variable ratio of 332:72 is less than 10:1, at 4.6:1.

The combined EFA demonstrated a reasonably clean seven factor structure with high KMO and 55% of the total variance explained by this model. The relabelled factors of Voice, Cough, Gastrointestinal Symptoms (GI), Airway Symptoms and Dysphagia, Throat Clearing, Lump in Throat and Life Events showed high internal correlation between the items within each factor. The removal of any items from the factors would not have offered improved correlation of items. The first three factors had adequately high mean loading scores. The latter four factors had lower mean loading scores. It is not surprising that the first three factors appeared strong, given the number of items dedicated to the symptoms of voice, cough and GI within the questionnaires. It is interesting to compare the EFA results with the ranked symptoms for the RSI and CReSS in Chapter 2. Whereas a feeling of a lump in the throat and clearing were the highest ranked individual symptoms, these factors had less items and lower correlations and mean loading scores. Again, this may reflect the fact that fewer items relate to these symptoms, than for instance gastrointestinal symptoms or voice, within the LPR-HRQL.

The following discussion will compare the individual questionnaire EFAs with the published literature:

RSI

Cathcart et al[50] explored the responses of 227 patients completing the RSI: 134 with catarrh symptoms, 61 with voice symptoms and 32 with respiratory disease. Cronbach's alpha was 0.75 for all items. A factor analysis using Principal Components Analysis as the extraction method, with oblimin rotation and Kaiser normalisation gave three principal components with igenvalues greater than one, but the scree slope suggested a two factor solution was more

suitable. Oblimin rotation found the two-factor model to be a neater fit. The two factors found were:

- Breathing difficulties or choking, Coughing after eating or lying down, Troublesome cough, Difficulty in swallowing, Heartburn or indigestion (Cronbach's alpha = 0.75). 4 of these items had factor loading scores > 0.5.

- Throat clearing, Globus, Postnasal drip, and Hoarseness or voice disorder. Cronbach's alpha =0.57, increasing to 0.77 when hoarseness was dropped. 3 out of the 4 had > 0.5 loading scores.

Printza et al [49]sought to validate a Greek version of the RSI using data from 53 patients and 172 control subjects. They included a 10^{th} item (throat pain) into the RSI. They followed the methods of Cathcart et al, to use oblimin rotation rather than orthogonal rotation. They also identified two factors in the RSI. They found the same two factors as Cathacrt et al, but found the globus item weakly loaded to the first factor too (0.27, but 0.62 on the second factor). No communality was noted below 0.35. Interestingly whereas hoarseness had been weakly loaded to the second factor in Cathcart et al's study, it was the top item in the second factor (0.92 loading). Both these studies were of sufficient size with a subject to variable ratio of >10:1.

Within the TOPPITS population, the RSI appeared the least appropriate data on which to conduct an EFA. Assuming an acceptable KMO score of 0.6, this data was just acceptable with a score of 0.61. Only 41% of variance was explained which reduced to 31% when the factor extraction method was changed from eigenvalues greater than one to a fixed two factor model. The three-factor model was not clean. The two-factor model was not cleaned in detail, but suggested:

- Troublesome Cough, Cough after eating or lying down, Throat clearing, Throat mucus or post-nasal drip, Hoarseness.

- Difficulty swallowing, Breathing difficulties, Lump in Throat.

Breathing difficulties cross loaded with similar loading values, and heartburn did not load onto either factor. Whilst there are some similarities between the present RSI EFA and the two published EFAs, given only nine items exist in the RSI, concordance between the analyses cannot be concluded. There does not appear to be a clear and reproducible factor structure to the RSI.

CReSS EFA

Interestingly, the EFA for the CReSS with factors defined by eigenvalues greater than one produced a similar factor model to the combined questionnaires EFA. However, with only one item related to voice within the CReSS, the individual CReSS EFA was not going to produce a Voice factor like the combined model. However, the first CReSS EFA did give a reasonably clean factor structure.

Drinnan et al [45] analysed the CReSS scores for 639 patients: 103 controls (staff, friends and hearing aid clinic attenders), 359 patients attending for oesophagogastroduodenoscopy (OGD), and 177 Ear Nose and Throat patients referred with throat symptoms. They conducted an EFA using principal components analysis as the extraction method, with varimax rotation (this is orthogonal rotation, but may be more appropriate as the CReSS has a wider range of questions, which may not correlate, and the factors may not be correlated as expected for the RSI). The subject to variable ratio was 639:34, and therefore >10:1. The CReSS had three statistically robust symptom factors: gastrointestinal (oesophageal); an upper airway factor – relating to cough, breathing, mucus and hoarseness; and a third, obstruction / choking globus factor (pharyngeal or throat)[45].

When the TOPPITS CReSS EFA was fixed to three factors, a clean structure was obtained with omission of only several items, with a high KMO of 0.89. A very similar factor structure was identified to that proposed by Drinnan et al[45]. The items within each factor are compared between TOPPITS and the previously published EFA by Drinnan et al in Table 4-10. The 3 items that were within the pharyngeal factor in TOPPITS only were the three items with the lowest loading scores.

Table 4-10 CReSS Factor Items within TOPPITS and Drinnan et al

CReSS Factors	Items common to both TOPPITS and Drinnan et al	Items only in Drinnan et al factor model	Items only in TOPPITS factor model
Oesophageal Factor	Heartburn Flatulence Regurgitation Acid/sour taste in mouth Gurgling Nausea Bloating Belching Feeling full too early in a meal Indigestion Stomach acid	Vomiting Pressure in the chest Low appetite Back pain Headache Bad breath	Rush of saliva
Upper Airway Factor	Throat clearing Excess mucus Mucus drip Coughing when upright Coughing after eating Coughing when lying down Wheezing Difficulty breathing Hoarseness.	None	None
Pharyngeal Factor	Lump in the throat Swallowing food Swallowing liquid Throat pain Feeling of things stuck in throat		Decreased appetite Choking Headache.

LPR-HRQL EFA

The single paper to assess the validity of this outcome measure analysed the results of 117 patients treated with proton-pump inhibitors[48]. Of the three questionnaires, this is the only EFA to be conducted solely on patients with throat symptoms. The specific methodological details of the factor analysis are not reported, stating a factor analysis of all (five) domains. Each domain had one single factor structure, except the voice / hoarse domain. Its second factor with an eigenvalue slightly greater than 1 was the reverse scoring item. The subject to variable ratio was 117:43 which is appreciably lower than the 10:1 rule of thumb, and less than 3:1. The overall sample size of 117 would be considered low for the number of observed variables.

Conducting an EFA on the TOPPITS LPR-HRQL again, materially adds to the evidence base on the factor structure of this specific outcome measure, and offers a significantly greater population size to draw conclusions from. Whilst the ratio of 332 patients to 43 items is not at the 10:1 level suggested earlier, only 29 of the 43 items were included in the present EFA. Only the voice, cough, throat and swallow domain items (29 in total) were included from the LPR-HRQL. Whilst the "overall impact of acid" items were included in the original construct validity work[57], it was felt that these items were too varied in their nature to reasonably comprise one set of related questions. The overall thermometer questions that follow each domain were similarly not included. The ratio of subjects to variables is therefore >10:1.The initial description by Carrau et al[57] did include the "overall impact of acid" items in the factor analysis. However, these were not included in the present EFA given the wide variety of quality of life aspects covered in these questions.

The present EFA showed similar results to Carrau et al, with a single factor for each domain of questions, other than the Throat domain. There were some items that cross-loaded on to two factors with all but one of these having a strong loading score in their domain factor and a weaker cross loading score in the other. The one obvious cross loading item was Throat24 – "The Need To Clear My Throat Makes It Hard To Talk". This item does clearly cover both throat clearing and voice, and it could be reassuring in terms of the EFA methodology that this item cross-loaded to both the voice and throat domain factors.

The Throat domain remained split over two factors, even with Throat24 removed. "People Notice How Much I Have To Clear My Throat", "The Need To Clear My Throat Makes It Hard To Talk", "I Feel Frustrated About Having To Clear My Throat So Often" comprised one factor, and the other: "Clearing My Throat Has A Negative Effect On Sex", "Clearing My

Throat Has A Negative Effect On Friendships", "I Avoid Social Events Because Of The Need To Clear My Throat". The latter three items formed the "life events" factor in the combined questionnaire EFA. It is interesting to observe that these three items loaded together in both the combined EFA and the individual LPR-HRQL EFA.

Following the CReSS EFA which demonstrated a reproducible 3 factor model, a similar 3 factor model applied to all three questionnaires did give an appropriate model. However, it was clear that the CReSS item related to a hoarse voice separated from the other CReSS items. The inclusion of the LPR-HRQL domains, with its 12 voice related items, will certainly have influenced the strength of the hoarse voice factor. However, it is worth considering that the CReSS questionnaire containing 34 items may be deficient in voice related symptoms.

4.5 Conclusions

Chapter 4 has demonstrated a clean seven-factor structure model for an exploratory factor analysis on the combined data from the RSI, CReSS and LPR-HRQL. The dimension reduction produced through this methodology could be useful to help define categories of persistent throat symptoms. The seven factors are clinically meaningful and relevant. As would be expected, the factors do categorise the main symptom groups from the questionnaires. These seven factors will inform the cluster analysis in Chapter 5. EFA and cluster analysis differ, in that EFA groups variables together, whereas cluster analysis groups together patients.

Exploratory Factor Analysis of the individual questionnaires gave poor models for the RSI, which were not in agreement with previously published data. The factor structure for the CReSS was clean and reproduced the Oesophageal, Upper Airway and Pharyngeal factors previously described within this questionnaire. The LPR-HRQL domains do contain one factor each, other than the throat domain which suffers with an item covering both voice and throat clearing and three questions covering life events which consistently load together as a separate factor.

Chapter 5 Cluster Analysis

5.1 Introduction

Chapter 4 demonstrated that seven symptom factors emerged from the TOPPITS baseline data with individual factors labelled as Voice, Cough, Gastrointestinal Symptoms (GI), Airway Symptoms and Dysphagia, Throat Clearing, Lump in the Throat and Life Events. This framework provides a basis for how persistent throat symptoms may be grouped into categories. Very little is known about how patients report related throat symptoms. A greater understanding or a clinically meaningful classification of related symptoms could aid improved management of these conditions. Chapter 5 will explore whether clusters of patients can be reliably defined according to their presenting symptoms. Exploratory factor analysis and cluster analysis differ in that EFA groups related variables together into factors, and cluster analysis groups together patients.

It is unlikely that a future study will utilise the RSI, CReSS and LPR-HRQL together. However, the wealth of symptom data that the three questionnaires has provided from TOPPITS could offer greater insights into patient symptom presentations. As in Chapter 4, the data from the three questionnaires will initially be analysed together. Cluster analysis investigates whether a population consists of distinct groups. Patients in one cluster will be similar, in this case on the basis of their throat symptoms, and will differ from patients in the other clusters. The seven factors from the exploratory factor analysis will be used as the variables to define the patients within the cluster analysis, rather than using all the individual symptom items from the three questionnaires. Cluster analysis of all the CReSS items was investigated by Papakonstantinou et al, but no distinct clusters of symptoms emerged[44]. A defined set of clusters of patients with similar symptom reporting could offer a clinically meaningful classification for the population of patients with persistent throat symptoms.

5.2 Methods

Cluster analysis techniques group, in this case, patients by assessing how similar they are in terms of measured variables. The distance between patients' scores on the variable of interest is measured. The smaller the number, the more similar the patients in terms of reporting that particular symptom. In its most simple form, taking the first patient, the first cluster can be made with another patient whose symptom score is closest to the first patient. Thereafter patients can be join the cluster through linkage techniques: the next patient is similar to the last, or the next is similar to either of the patients in the cluster, or the next is similar to the average similarity of the patients in the cluster[71]. Another technique is known as Ward's

method, which enters patients into clusters to minimise the variance of the measured variables within each cluster[71].

The baseline data were analysed using three cluster analysis methods: Two-Step cluster analysis, K-means cluster analysis and the Hierarchical method. Of the 3 options, K-means analysis is a non-hierarchical method, there is a specific hierarchical method, and the Two-step analysis is a mix of both non-hierarchical and hierarchical methods. Non-hierarchical methods, in which the relationship between clusters is undetermined, would seem appropriate. However, hierarchical methods are more frequently used in this type of analysis and were used in the previous CReSS cluster analysis[44].

The factor score variables included in the cluster analysis were incorporated as standardised values, produced from the exploratory factor analysis in Chapter 4. Using standardised values allows the results to be interpreted with ease and ensures all variables contribute evenly to a scale; Standardised factor scores were computed such that they had mean of zero and standard deviation of one. The value of each score is a measure of the spread from the group mean, as a proportion of the standard deviation. Cluster analysis was performed for the three methods initially using all seven factors: Voice, Cough, Gastrointestinal Symptoms (GI), Airway Symptoms and Dysphagia, Throat Clearing, Lump in the Throat and Life Events. Cluster analysis was performed then with all or a reduced number of factors to establish if a reproducible model could be defined over the three methods.

The number of clusters can either be left open, and automatically defined by the statistical analysis, or be specified before the analysis is performed.

5.2.1 *Two-step Cluster Analysis*

This method initially performs a pre-clustering step to identify groups before running a hierarchical clustering algorithm. The log-likelihood distance measure was used, which assumes the variables are normally distributed and independent. The distribution of the standardised scores was assessed using histograms with overlying normal curves, alongside the skewness and Kurtosis scores. Transformation of the standardised scores was explored if they were not normally distributed, to ascertain if the distribution of the standardised scores affected the two-step cluster analysis results.

When the number of clusters was left open, to be automatically determined by the model, both Schwartz's Bayesian Criterion and Akaike's Information Criteria methods were explored.

Several graphical outputs are produced in SPSS:

The Model Viewer produced gives a model summary for the analysis. This gives an indicator of cluster quality on a scale of poor to fair to good. This is based on the degree of cohesion of the clusters and the silhouette, i.e. the extent to which the clusters overlap or are separate. It also gives a pie chart of cluster sizes and a ratio of the largest cluster to smallest cluster; less than three is desirable[72].

The Predictor Importance output demonstrates how much each variable determines the cluster model, and is important in assessing if any strong variables exist that entirely explain the cluster model (such as sex).

The Cluster Comparison output plots the cluster median and interquartile range for each variable, overlying the overall population median and interquartile range. The advantage of the Two-step analysis over the K-means analysis is the ability to assess easily the range of values around the median score for each cluster variable. The cluster membership values can be saved as a data output for each patient within SPSS.

To validate the number of clusters in a two-step cluster analysis, two methods are recommended[73]. Firstly, assessing the cluster quality scale between different models that vary the number of clusters to be defined. Secondly, through an analysis with automatically determined cluster numbers either a Schwartz's Bayesian Criterion (SBC) an Akaike's Information Criteria (AIC) line graph can be plotted. The pivot table option is chosen, the appropriate SBC or AIC values within the table highlighted and a line graph selected. Where the line reaches its lowest point depicts the ideal number of clusters to be specified. AIC methods are considered to produce more complex outcomes than Bayesian Criterion methods.

5.2.2 *K-means Cluster Analysis*

Cluster membership was presented graphically using bar charts and the contribution of each variable to the cluster solution analysed using analysis of variance (ANOVA). Variables that were not significant on ANOVA, and therefore not contributing to a patient's cluster membership, were removed. The number of suitable clusters in the data was expected to be between two to five clusters. The number of appropriate clusters for a defined cluster number analysis followed the SBC or AIC line graph explored with the Two-step analysis techniques. Informally, the K-means model was assessed using the ANOVA table for F values and significance, the size of the cluster groups, and the number of iterations required for an analysis to converge (should be less than 10 iterations)[74].

5.2.3 *Hierarchical Cluster Analysis*

A dendrogram plot was produced. The Wards' cluster method was used; to produce similar sized clusters. It is also an appropriate method for data that has no outliers. Given the Likert scale of all the items within the questionnaire, outliers cannot exist - as would be expected with a continuous scale such as BMI. The squared Euclidean distance was used for interval data, which is the square root of the sum of squared distances between two observations.

Given the large dataset and the anticipated large dendrogram, precluding inclusion in this thesis, the cluster membership number for each patient was used to generate descriptive statistics and bar charts of the symptom groups. 95% confidence intervals for the mean factor scores within clusters were calculated according to the group mean +/- 1.96 x sd / $\sqrt{n}$.

5.2.4 *Comparisons between the cluster analysis techniques*

The three cluster analysis techniques give differing outputs through SPSS. To allow an assessment of comparison amongst the techniques a standard output was required. For each methodological analysis the cluster membership group for each patient was retained. Box and whisker plots were generated to depict the standardised factor scores amongst the cluster groups for the three techniques. The box and whisker plots depict the median and interquartile range as the box, with the whiskers showing the highest and lowest scores – but which are restricted to within 1.5 times the interquartile range. Scores outside that range are shown as outliers.

5.2.5 *Co-variate analysis amongst clusters*

The cluster membership values from each analysis technique were saved as data outputs for each patient within SPSS. Definable clusters were plotted graphically against the covariates of age and BMI. Scatterplots and box and whisker plots were drawn up to include cluster membership, to assess if these covariates differed between symptom cluster groups.

5.2.6 *Cluster analysis of RSI data alone*

The seven factors from the EFA are similar to the structure of the RSI questionnaire. Therefore cluster analysis of the most widely used questionnaire, the RSI, was performed to ascertain if the findings differ from the above analysis using the defined factors, and also to provide a comparison for any future studies. The methodology was identical to that described in 5.2.1.

5.3 Results

Two clusters were identified when each of the three cluster analysis techniques were used and it was specified that the number of clusters should be determined automatically (i.e. number of clusters undefined) There was a cluster who scored below the mean or median for all seven factor standardised scores and a cluster who scored above the mean or median. This was consistent if the number of factor scores included in the analysis was reduced. The results for the Two-step analysis and K-means analysis are represented in the Appendices: Figure 7-11, Figure 7-12, Figure 7-13, Figure 7-14 and Table 7-6. A two cluster model was not explored for the hierarchical methods given these results. The two cluster model was inappropriate to draw any further conclusions from.

To determine how many clusters should be defined in the subsequent analyses, both an AIC line graph and a line graph to depict SBC were plotted from data derived from a two-step cluster analysis of all seven factor scores (Figure 5-1, Figure 5-2). For the AIC graph, the line is steepest to two clusters, and shallows after each additional cluster. However its trough appears to be around seven clusters. The SBC line graph is steepest to two, descends further to three and is near flat to four, before rising again. Three clusters would appear therefore the most appropriate model to explore, with consideration of a four cluster model too. Seven clusters was considered too many to offer a clinically meaningful model.

The number of factors to include in the cluster analyses was explored. The results for all seven factors is presented. Exploring fewer factors, prioritising those with the highest mean factor loading scores from the exploratory factor analysis, suggested that the three highest scored factors: Voice, Cough and Gastrointestinal symptoms may offer a clinically meaningful model. The results present these three factor cluster models.

Figure 5-1 Line Graph showing Akaike's Information Criteria, derived from a two-step cluster analysis of seven factor scores

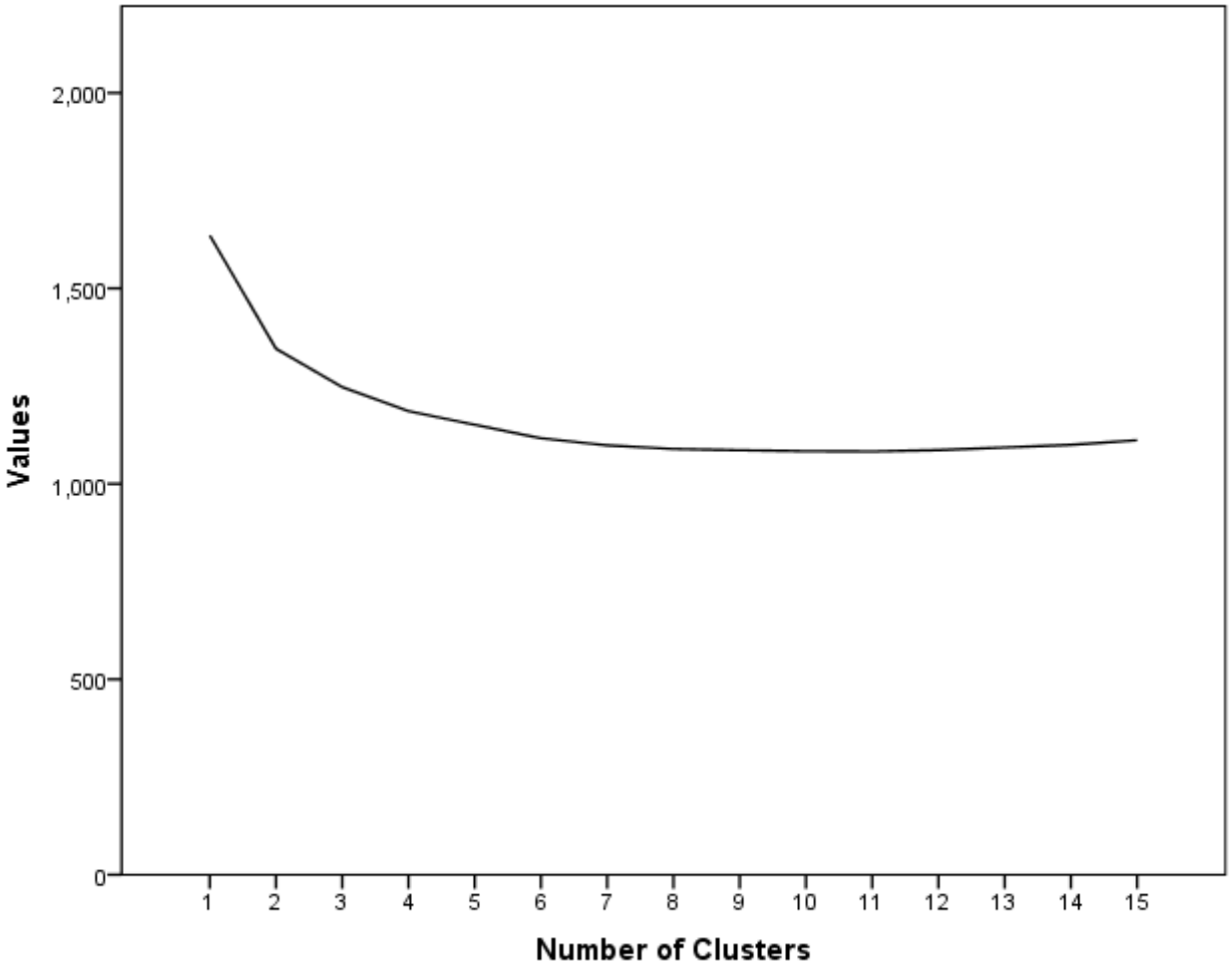
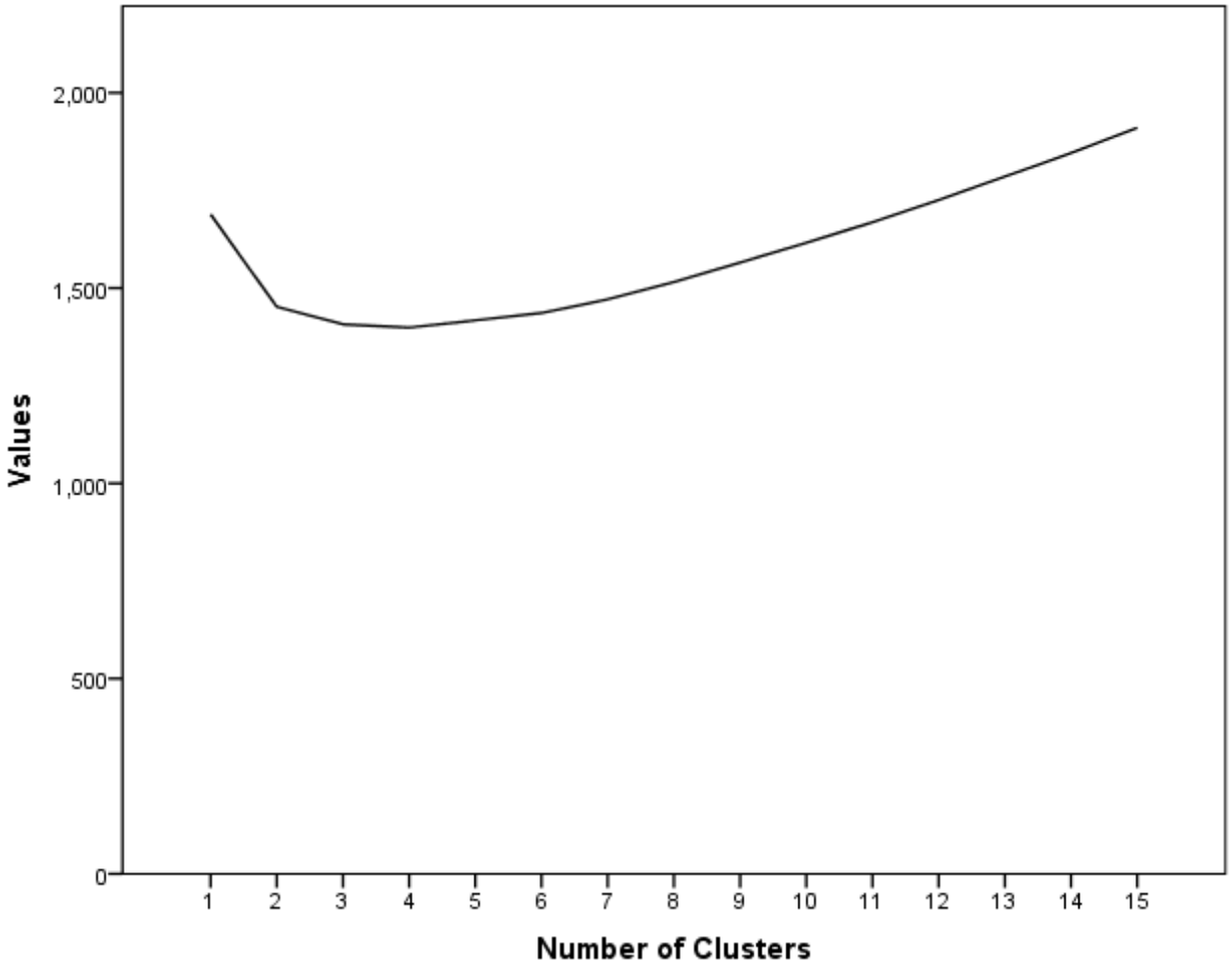


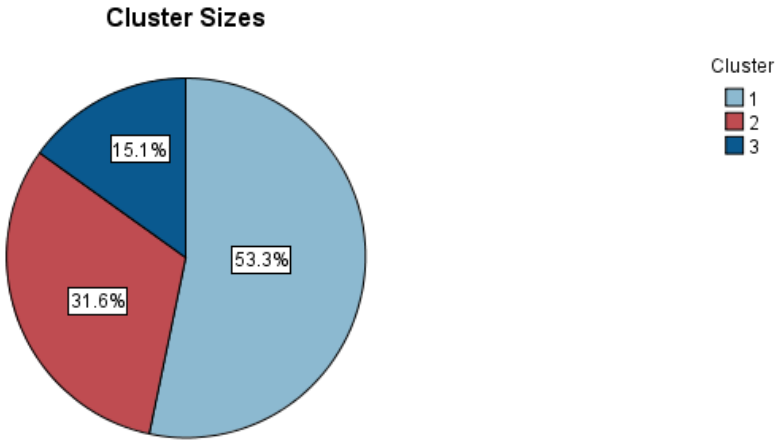
Figure 5-2 Line Graph showing Schwarz's Bayesian Criteria, derived from a two-step cluster analysis of seven factor scores



5.3.1 *Two-Step Cluster Analysis with three clusters specified and seven factors included*

The two-step cluster analysis methodology was repeated with three clusters specified rather than the number of clusters left open. Figure 5-3, Figure 5-4 and Figure 5-5 show the results outputs: This showed a poor to fair model. Voice was the strongest predictor of this model, with lump in the throat being the weakest predictor. However, six of the seven factors contributed significantly to the model; there was no single factor which dominated and influenced the model over the other factors. Cluster one scored below the TOPPITS population median across all seven factors, cluster two scored above the median, and cluster three scored above the median with several scores lying above the upper quartile for the whole population. On closer inspection, cluster two had higher relative GI scores and cluster three had relatively higher voice, life events and cough scores.

Figure 5-3 Cluster size: Two step cluster analysis for seven factor model with three clusters specified



Size of Smallest Cluster	50 (15.1%)
Size of Largest Cluster	177 (53.3%)
Ratio of Sizes: Largest Cluster to Smallest Cluster	3.54

Figure 5-4 Predictor importance: Two step cluster analysis for seven factor model with three clusters specified

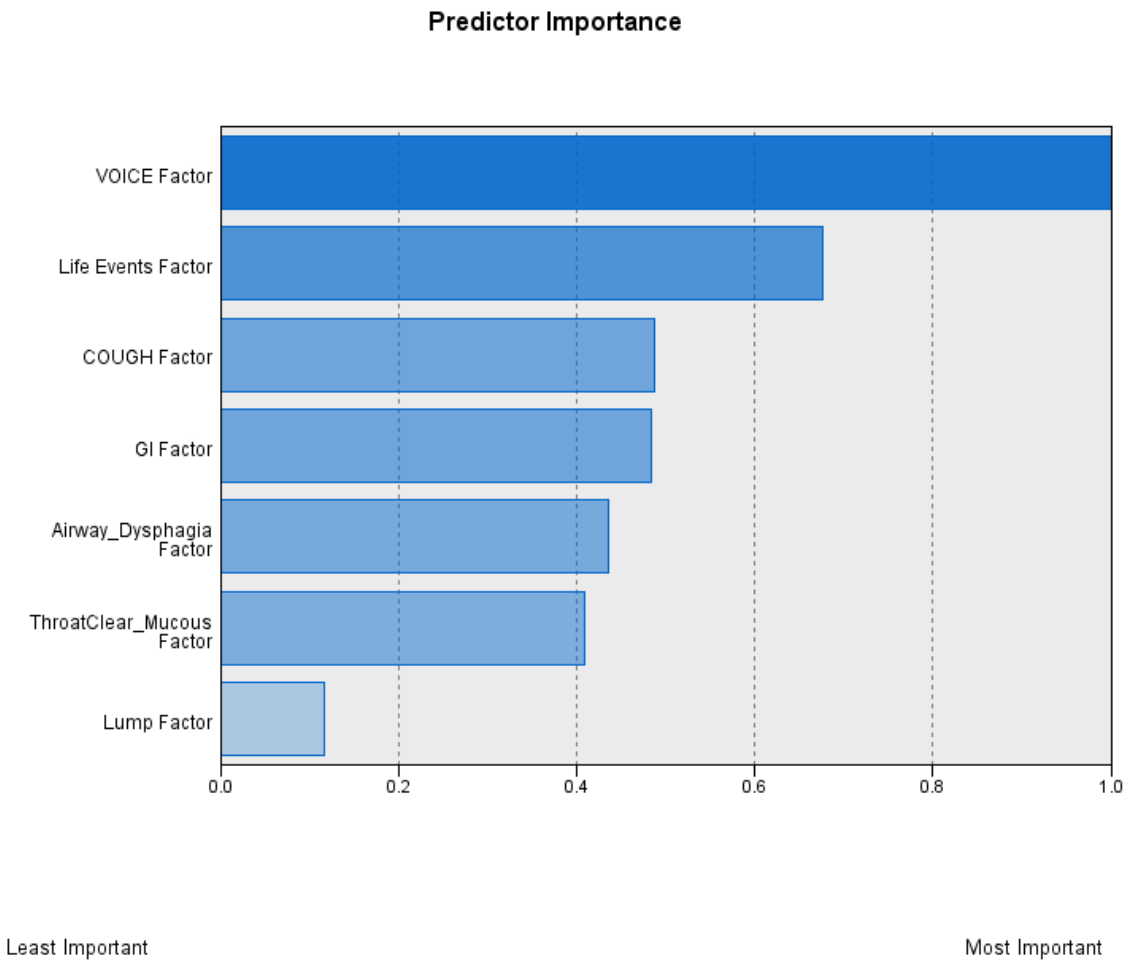
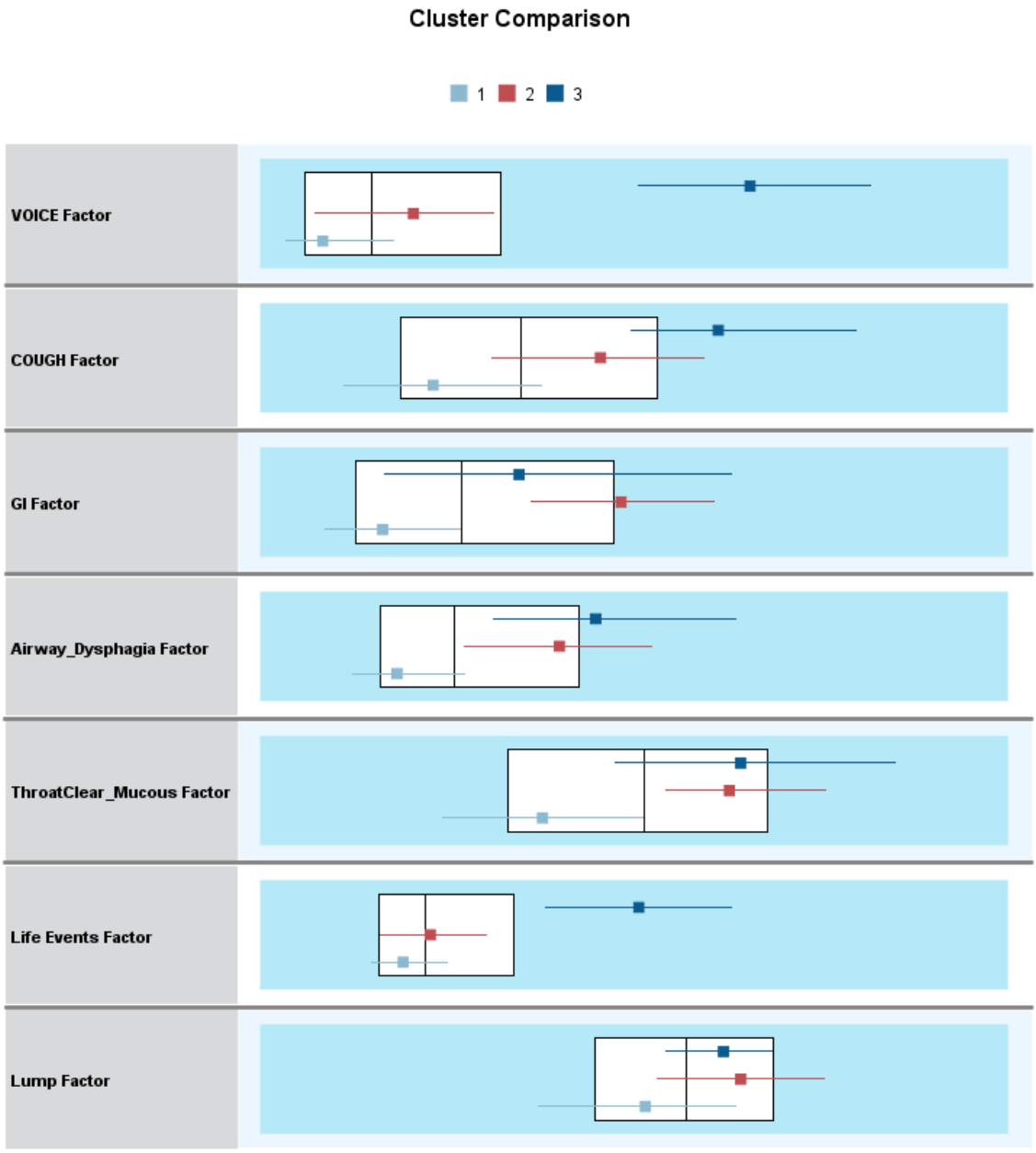


Figure 5-5 Two step cluster analysis: seven factors, three clusters specified



The white boxes represent the median and interquartile range for the population. The lines and square dot represent the median and IQR for the cluster. The medians differ to the mean of 0 for each factor.

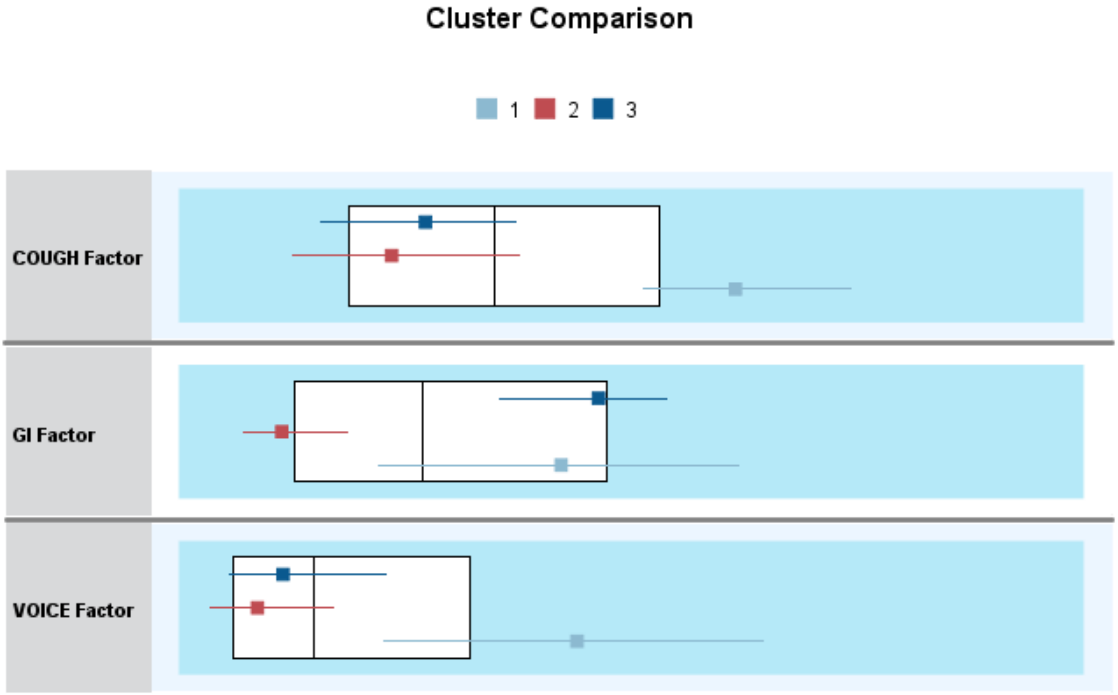
5.3.1.1 Two-step cluster analysis with 3 clusters specified and the 3 highest factors included

Reducing the number of factors in the cluster analysis model was explored, resulting in the three highest loading mean factors: voice, cough and GI symptoms entered into the two –step cluster analysis. This model gave a poor to fair cluster quality (Figure 7-15), but with equal

sized clusters (Figure 7-16), and the three factors were even in their strength of prediction within the model (Figure 7-17). If the Euclidean Distances measure was used instead of the Log-Likelihood measure, the three clusters were unequal in distribution, with one very large cluster and two smaller clusters. This option was not considered any further. The results outputs are shown in Figure 5-6. Cluster one of 103 patients scored high across the three symptoms. All three symptom groups were scored above the population median, but relatively higher scores were seen in the cough and voice symptoms than in the GI symptoms. Cluster two of 134 patients scored below the population median for all 3 symptom groups. Cluster three of 95 patients scored below the population median for cough and voice, but above the median for GI symptoms.

The standardised factor scores are displayed in histograms and overlying normal curves in Table 7-7. Not all factors were normally distributed. The standardised scores were transformed using a log transformation +8, to ensure all scores were positive. The two-step cluster analysis was repeated using these transformed scores. The results did not differ using the transformed standardised scores.

Figure 5-6 Two-step cluster analysis with three factors, three clusters specified



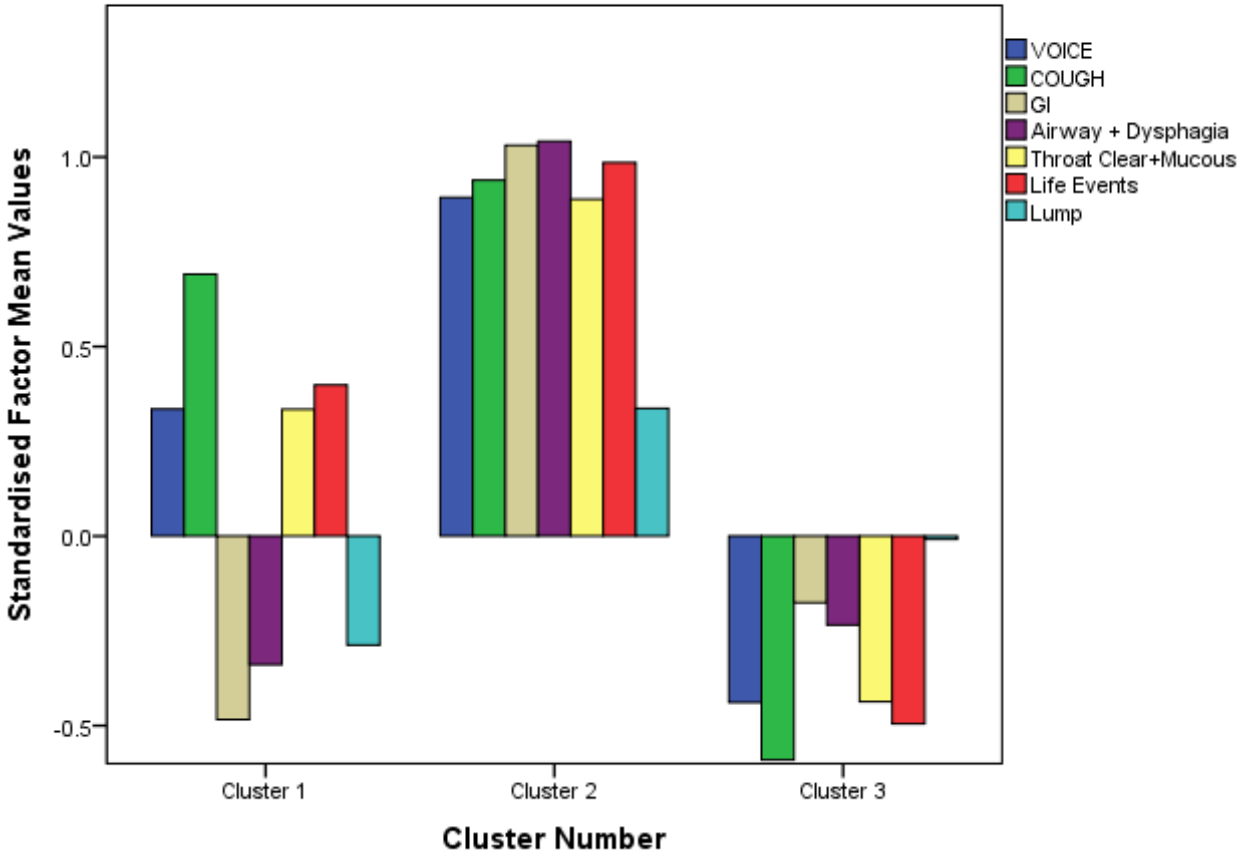
5.3.1.2 Two-step cluster analysis with three highest factors / symptom groups, four clusters specified.

Figure 7-18 in the appendix shows the model summary for this analysis. The cluster quality was particularly poor, which precluded any further cluster analysis or conclusions to be drawn from this model.

5.3.2 *K-Means Cluster Analysis with 3 clusters specified*

The seven-factor K-means cluster analysis methodology was repeated with three clusters specified. Figure 5-7 shows the 3 clusters: cluster one of 73 patients with variable scores across the seven symptom groups (voice, cough, throat clearing and life events greater than the TOPPITS population mean – GI symptoms, airway/ dysphagia and lump in throat lower than the population mean), cluster two with 67 patients generally scoring high across the symptom groups, and cluster three with 192 patients scoring generally low across the symptom groups. This solution converged in nine iterations The ANOVA table showed that all seven symptom groups significantly contributed to cluster membership, the Lump in Throat again demonstrating the lowest F value (see appendix Table 7-8).

Figure 5-7 Bar chart to show three cluster distribution for seven factors using K-means analysis



The K means method was repeated this time with three symptom groups using the highest mean factor loading scores from the EFA: Voice, Cough and GI, again - to assess if a reduced number of symptoms would offer more straightforward and clinically meaningful clusters. This is represented graphically in Figure 5-8. It suggests that cluster one, with 103 patients, may have higher reported GI symptoms than the TOPPITS population mean. Cluster two with 180 patients reported generally lower scores across the three symptom groups. Cluster three with 49 patients may have higher reported voice symptoms. This model did not converge after 10 iterations, hence the analysis may not be appropriate to draw conclusions from. A one-way ANOVA with post hoc Bonferoni analysis showed that the GI scores were not significantly different between cluster two and three. Table 5-1 shows that all 3 symptom groups had high F scores and significantly contributed to patients' cluster membership.

Figure 5-8 Bar chart showing three cluster distribution for three factors using K-means analysis

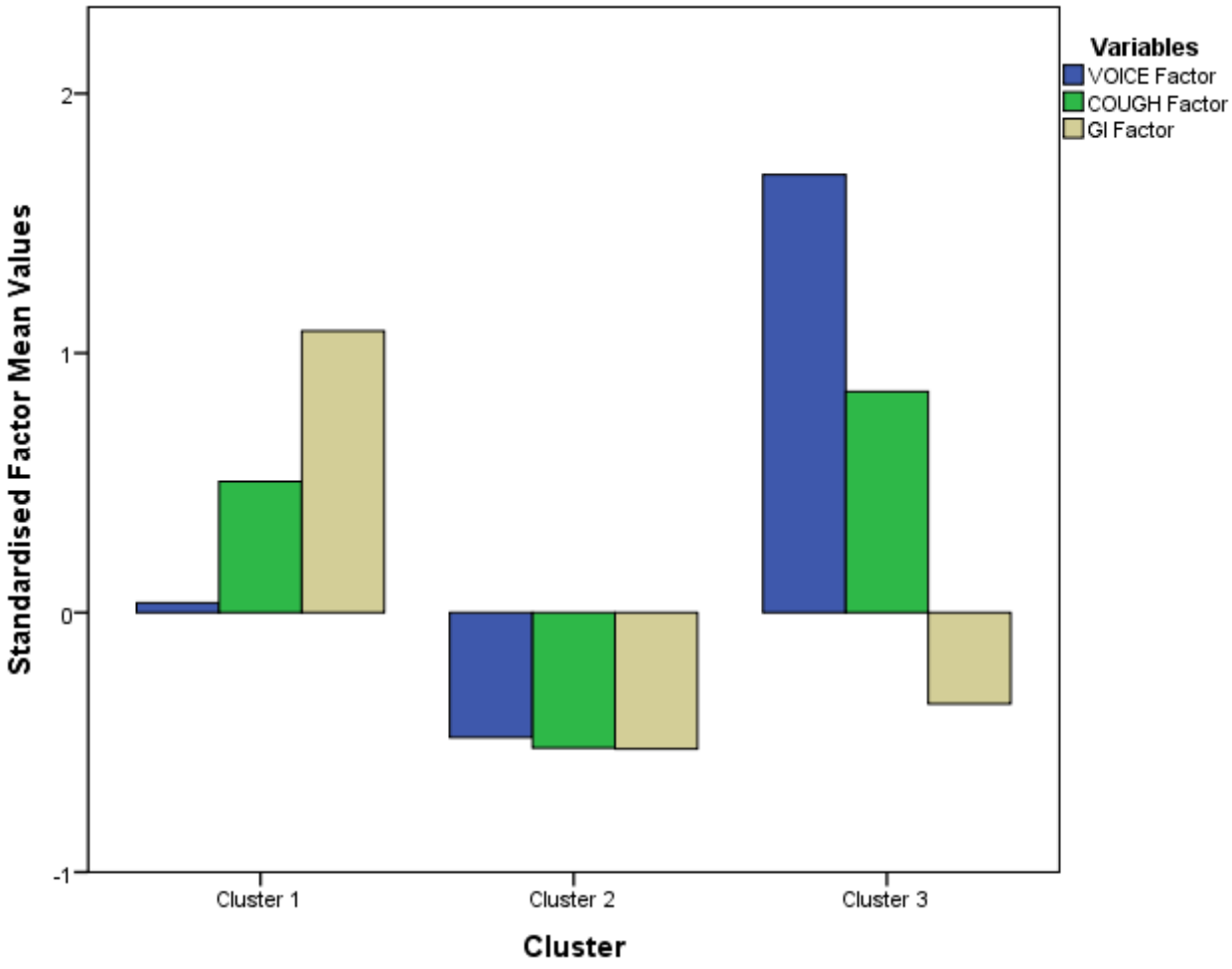


Table 5-1 ANOVA for K means three factor three cluster model

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
VOICE Factor	90.61	2	.41	329	221.70	p<0.001
COUGH Factor	55.34	2	.60	329	91.64	p<0.001
GI Factor	88.48	2	.39	329	226.94	p<0.001

df, degrees of freedom; F, F-score, Sig, statistical significance

5.3.3 *Hierarchical Cluster analysis*

5.3.3.1 Hierarchical cluster analysis with clusters specified for seven factors.

The dendrograms produced from this analysis are not included in the thesis, given their physical size to reproduce. When all seven symptom group factors were included, three clusters emerged of 131, 116 and 85 patients. Table 5-2, Table 5-3, Table 5-4 and Figure 5-9 show the raw data and the graphical representation of these 3 clusters. Cluster one scored below the overall TOPPITS population mean scores for all seven symptom groups. Cluster two scored above the TOPPITS population mean scores for GI symptoms, Throat clearing and mucus, and the lump symptom. Cluster three scored above the overall TOPPITS population mean scores for all seven symptom groups.

Table 5-2 Cluster one: hierarchical cluster analysis for seven factor model

	n=	Minimum	Maximum	Mean	Std. Dev.
VOICE Factor	131	-.99	1.41	-.42	.59
COUGH Factor	131	-1.58	1.33	-.60	.65
GI Factor	131	-1.36	1.84	-.53	.61
Airway_Dysphagia Factor	131	-1.51	2.67	-.19	.88
ThroatClear_Mucus Factor	131	-2.12	.46	-.83	.58
Life Events Factor	131	-1.33	.54	-.50	.35
Lump Factor	131	-2.79	2.16	-.19	.99

Std dev, standard deviation.

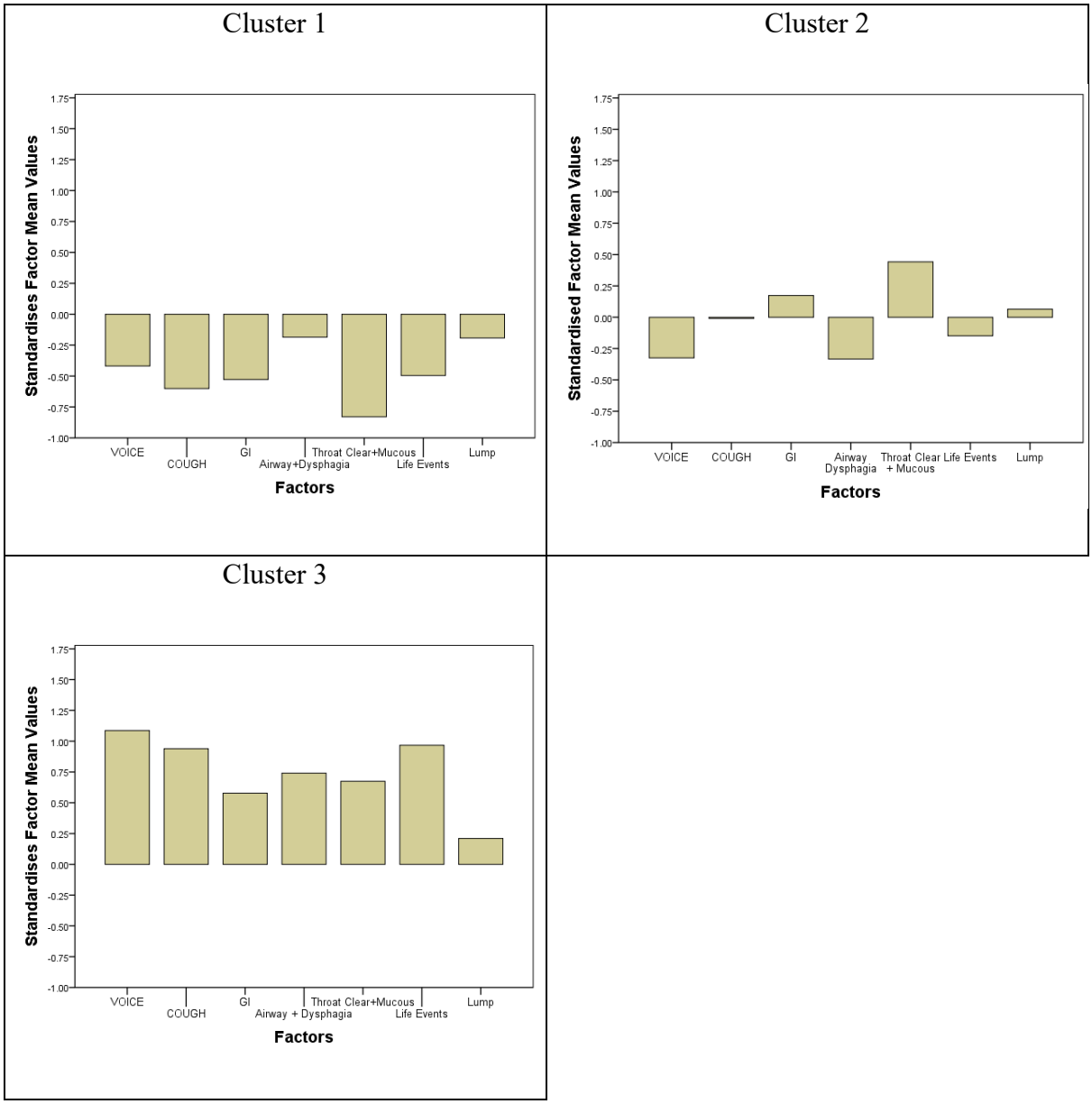
Table 5-3 Cluster two: hierarchical cluster analysis for seven factor model

	n=	Minimum	Maximum	Mean	Std. Dev.
VOICE Factor	116	-.94	1.81	-.32	.53
COUGH Factor	116	-1.45	1.98	-.01	.80
GI Factor	116	-1.35	2.19	.17	.87
Airway_Dysphagia Factor	116	-1.46	1.48	-.33	.62
ThroatClear_Mucus Factor	116	-1.78	1.76	.44	.69
Life Events Factor	116	-1.52	2.10	-.15	.67
Lump Factor	116	-1.93	2.14	.06	.77

Table 5-4 Cluster three: hierarchical cluster analysis for seven factor model

	n=	Minimum	Maximum	Mean	Std. Dev.
VOICE Factor	85	-.77	3.28	1.09	1.11
COUGH Factor	85	-1.06	2.61	.94	.84
GI Factor	85	-1.30	2.91	.58	1.10
Airway_Dysphagia Factor	85	-1.21	3.34	.74	.98
ThroatClear_Mucus Factor	85	-1.23	2.03	.67	.75
Life Events Factor	85	-.96	4.17	.97	1.07
Lump Factor	85	-1.96	1.74	.21	.82

Figure 5-9 Bar charts to show hierarchical cluster analysis results for three specified clusters, on seven factors



5.3.3.2 Hierarchical cluster analysis with three clusters specified for three factors.

When the 3 highest mean scored factors were included – as in the final Two-step and K means analysis including Voice, Cough and GI factors, three clusters emerged of 153, 117 and 62 patients. Table 5-5, Table 5-6, Table 5-7 and Figure 5-10 show the raw data and the graphical representation of these three clusters. Cluster one scored below the overall TOPPITS population mean scores for all three symptom groups. Cluster two scored above the TOPPITS population mean scores for GI symptoms, at around the population mean for Cough, and below the mean for Voice symptoms. Cluster three scored above the overall TOPPITS population mean scores for all three symptom groups. There would appear graphical trends towards lower GI scores than Voice and Cough in cluster one, and higher Voice scores than

Cough and GI symptoms in cluster three. In cluster one the 95% confidence interval for GI symptoms did not overlap with those of the voice factor and cough factor. Cluster two was clear graphically, and when observing the 95% confidence intervals, in having higher GI scores and lower voice scores. The voice factor score in cluster three was higher than the cough and GI factors, with non-overlapping 95% confidence intervals when observing the voice with the cough and GI factors.

Table 5-5 Cluster one: hierarchical cluster analysis for three factors

	N	Minimum	Maximum	Mean	Std. Dev.	95% CI
VOICE Factor	153	-.99	1.41	-.36	.58	-.45 to -.27
COUGH Factor	153	-1.54	1.98	-.36	.87	-.50 to -.22
GI Factor	153	-1.36	.27	-.77	.35	-.83 to -.72

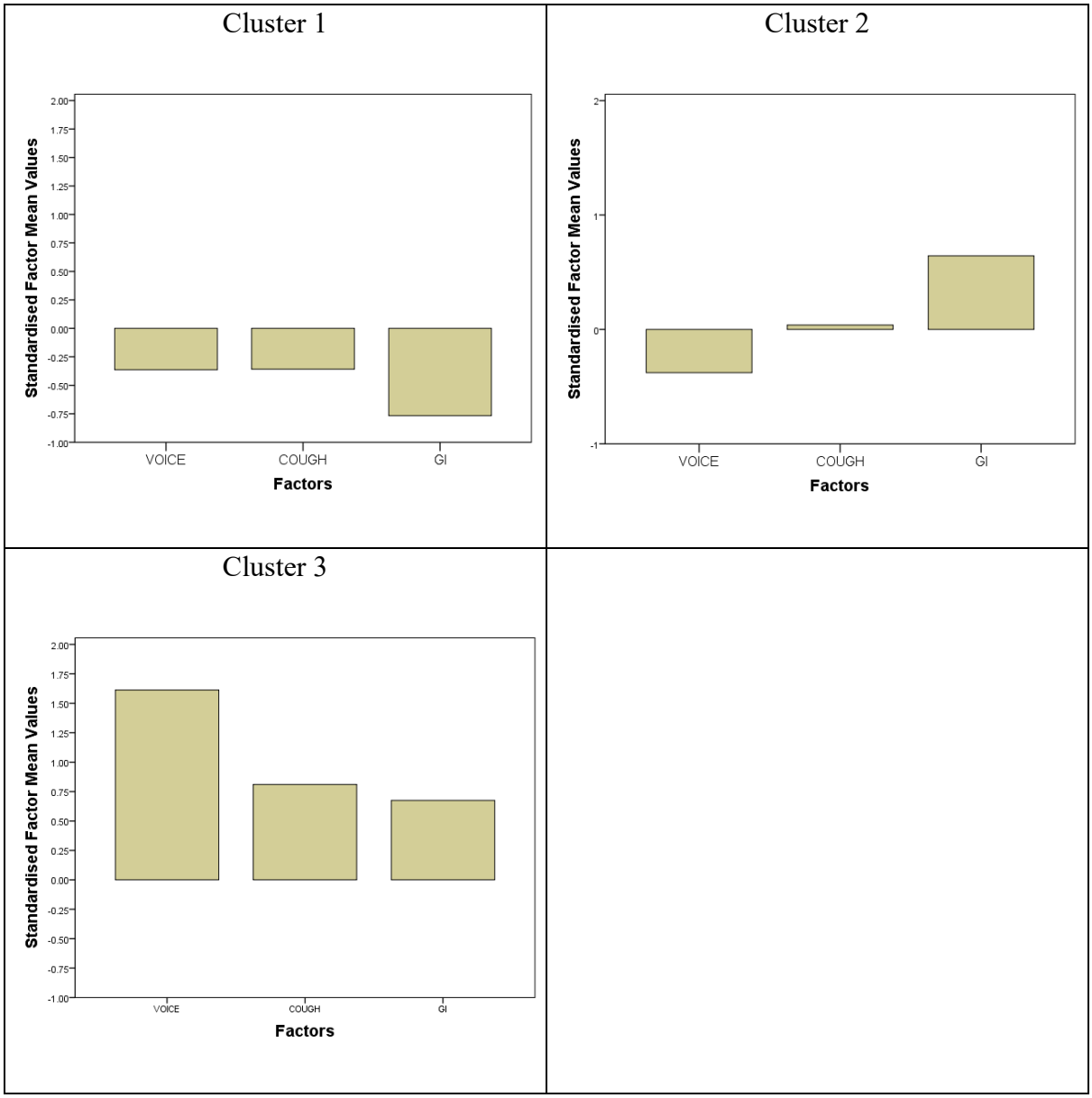
Table 5-6 Cluster two: hierarchical cluster analysis for three factors

	N	Minimum	Maximum	Mean	Std. Dev.	95% CIs
VOICE Factor	117	-.94	.52	-.38	.43	-.46 to -.30
COUGH Factor	117	-1.58	2.61	.04	.85	-.11 to .19
GI Factor	117	-.33	2.19	.64	.58	.54 to .75

Table 5-7 Cluster three: hierarchical cluster analysis for three factors

	N	Minimum	Maximum	Mean	Std. Dev.	95% CIs
VOICE Factor	62	-.04	3.28	1.61	.85	1.4 to 1.8
COUGH Factor	62	-1.06	2.59	.81	.89	.59 to 1.0
GI Factor	62	-1.28	2.91	.68	1.15	.39 to .97

Figure 5-10 Bar charts to show hierarchical cluster analysis results for three specified clusters, on three factors



5.3.4 *A comparison of the three cluster analysis techniques*

As shown above, the outputs from each cluster analysis technique differ markedly within SPSS. The number of patients in each cluster for each method is shown in Table 5-8. The cluster distributions are shown in Figure 5-11, Figure 5-12 and Figure 5-13 for the 3 cluster models of the three highest scored factors of Voice, Cough and GI.

Table 5-8 Number of patients in each cluster for each cluster analysis technique - three specified cluster, three factors

	Two-step cluster analysis	K-means cluster analysis	Hierarchical cluster analysis
Cluster 1	103	103	153
Cluster 2	134	180	117
Cluster 3	95	49	62

Figure 5-11 Box and whisker plot showing the Two-Step cluster analysis results for three factors and three specified clusters

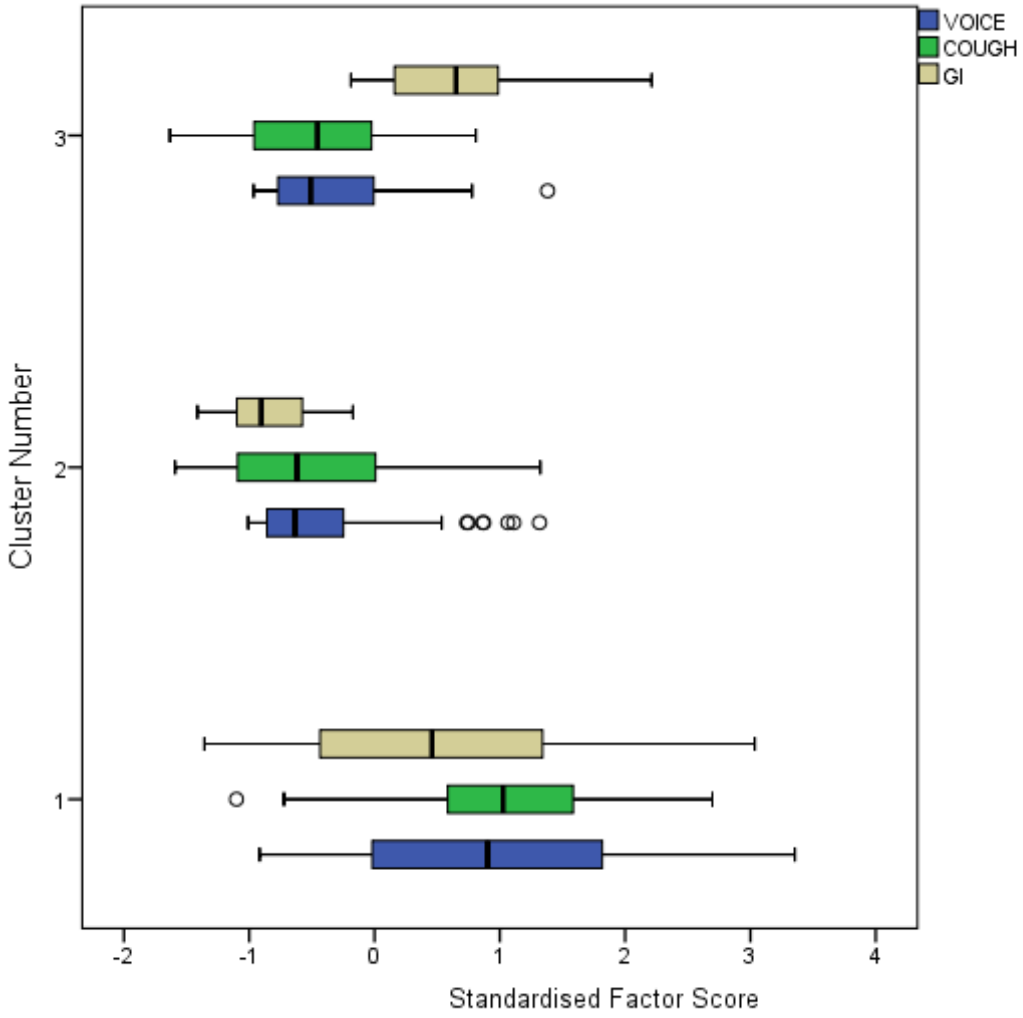


Figure 5-12 Box and whisker plot showing the K-means cluster analysis results for three factors and three specified clusters

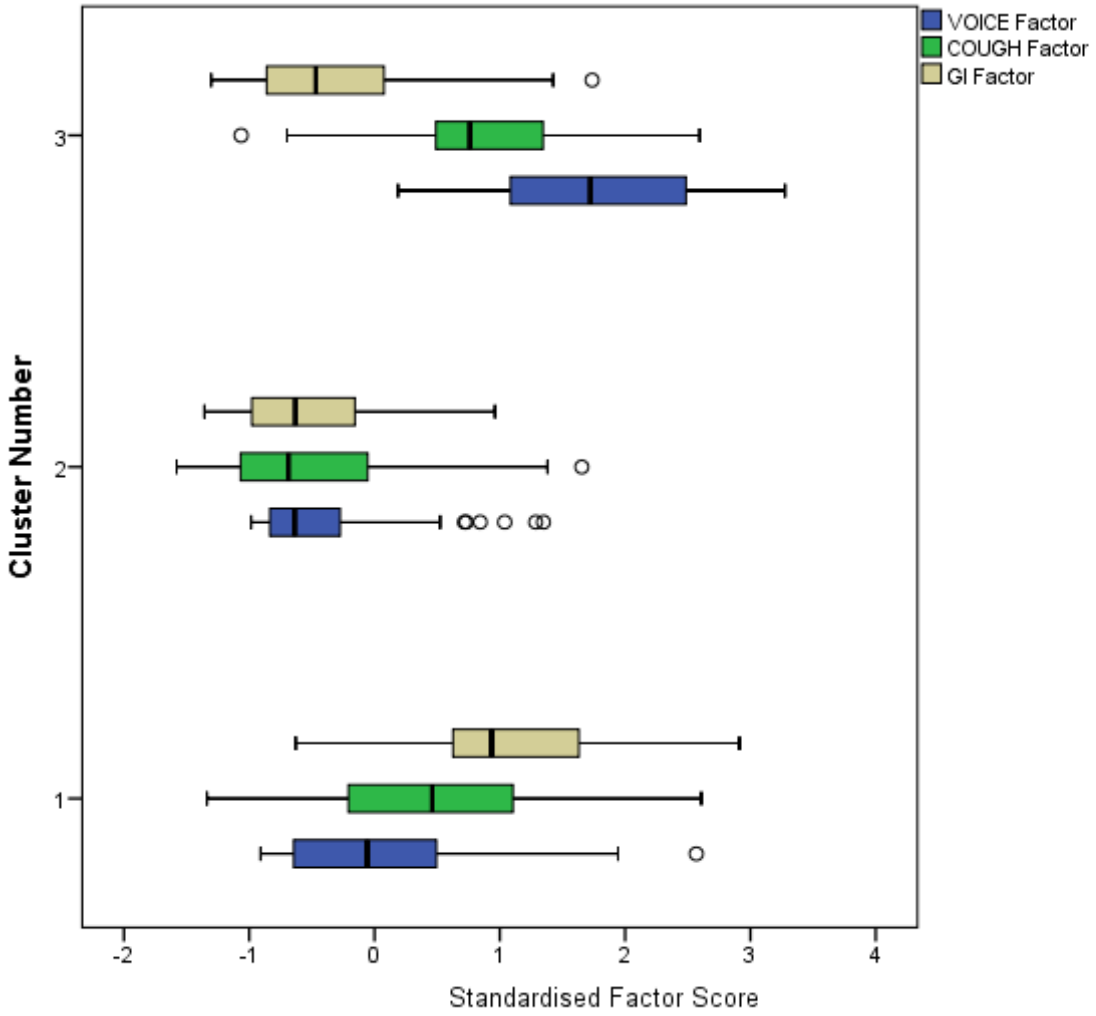
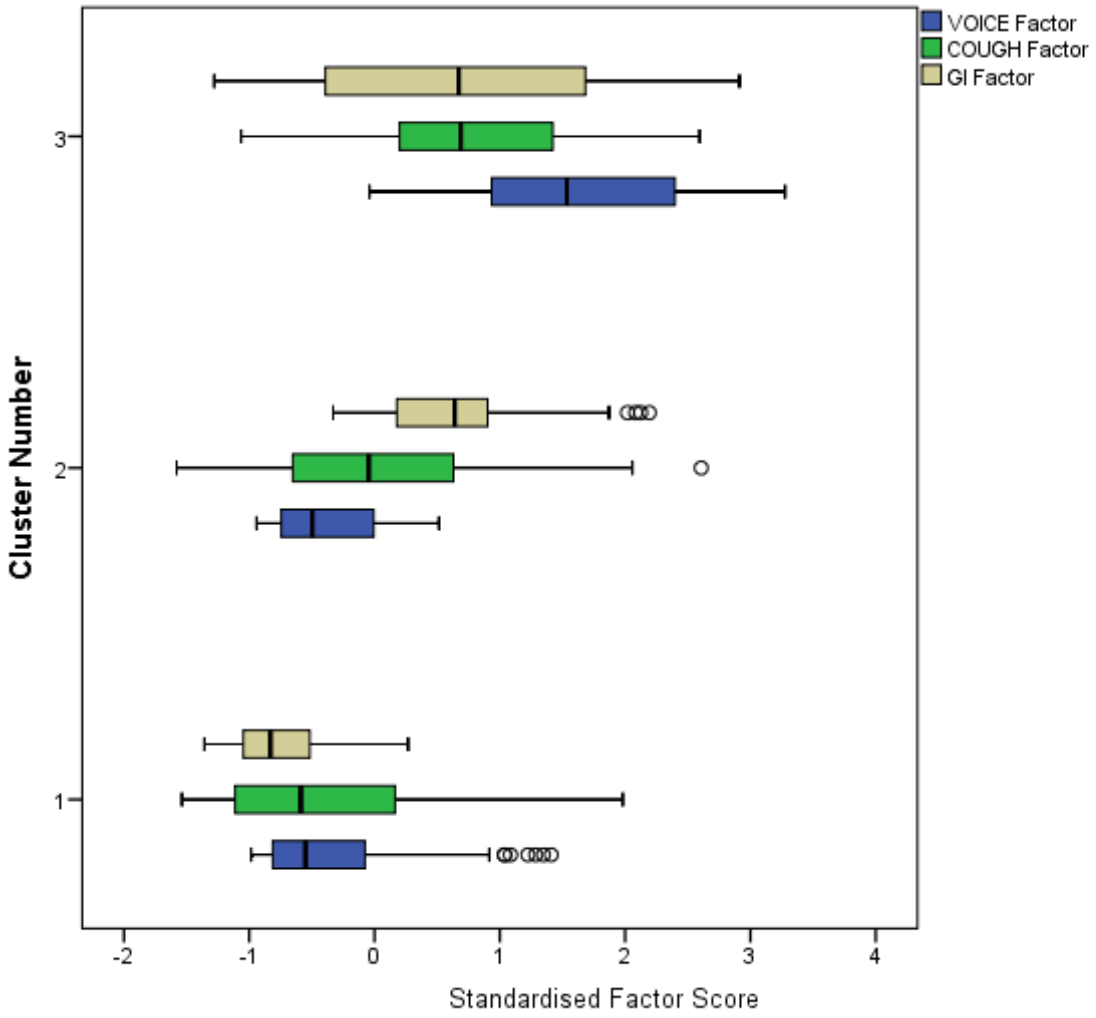


Figure 5-13 Box and whisker plot showing the hierarchical cluster analysis results for three factors and three specified clusters

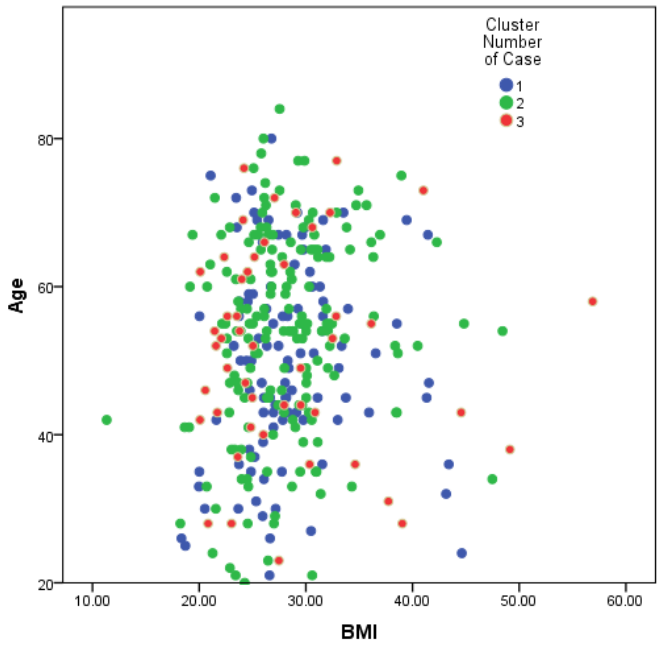


5.3.5 *Co-variate analysis amongst clusters*

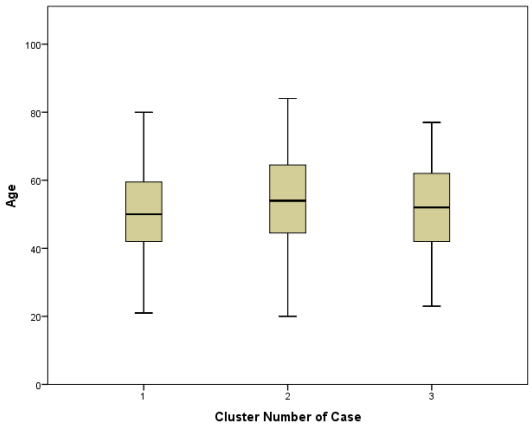
The three clusters defined by the K means analysis using the three factors of voice, cough and GI symptoms were saved with their cluster membership values. The relationship between the cluster and the demographics of patients within each cluster is graphically displayed in Figure 5-14. The scatterplot shows no discernible pattern of BMI or age across the three clusters. The median, interquartile range and range for the patients' age in the three clusters appear so similar across the clusters that formal analysis was not performed. Likewise, the distribution of BMI appears similar across the three clusters. Given the lack of any trends for differences in co-variates amongst clusters this analysis was not repeated following the Two-Step or Hierarchical cluster analysis techniques.

Figure 5-14 Exploring the relationship between cluster and demographics (age and BMI).

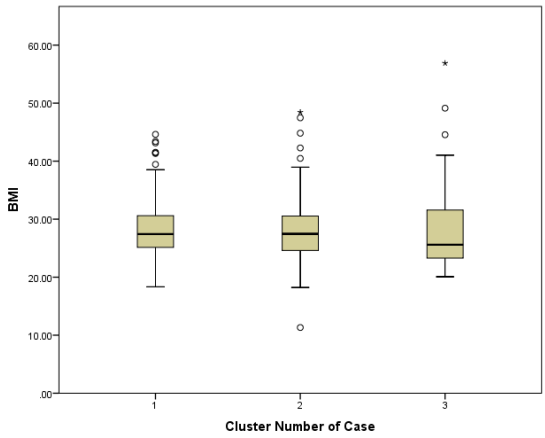
Scatterplot of BMI versus Age according to cluster membership groups



Box and Whisker plot of age across 3 cluster groups



Box and whisker plot of BMI across 3 cluster groups



5.3.6 *Reflux Symptom Index Questionnaire Cluster Analysis*

5.3.6.1 Two-step RSI cluster analysis

Line graphs of AIC and SBC are shown in Figure 5-15 and Figure 5-16. The AIC graph showed the curve shallows between 3 and 4 clusters. The SBC curve had a trough at three clusters. Three cluster were specified therefore for the analyses. Figure 7-19, Figure 7-20 and Figure 5-17 show the model outputs. The model had a poor cluster quality. Excess throat mucus was the greatest predictor of the model, with heartburn being the least strong predictor to the model. Cluster one of 85 patients scored at or above the population upper quartile for

all items except the heartburn item, which scored around the population median. Cluster two of 162 patients scored around the population upper quartile for excess mucus and throat clearing, around the population median for coughing after eating or lying, throat clearing, hoarse voice and heartburn, and around the lower quartile for breathing difficulties / choking, swallowing difficulties and troublesome cough. Cluster three of 85 patients scored below the population median for six items with swallowing difficulties, lump in the throat and breathing difficulties scored near the population median.

Figure 5-15 Line graph of Akaike's Information Criterion for RSI Two step cluster analysis

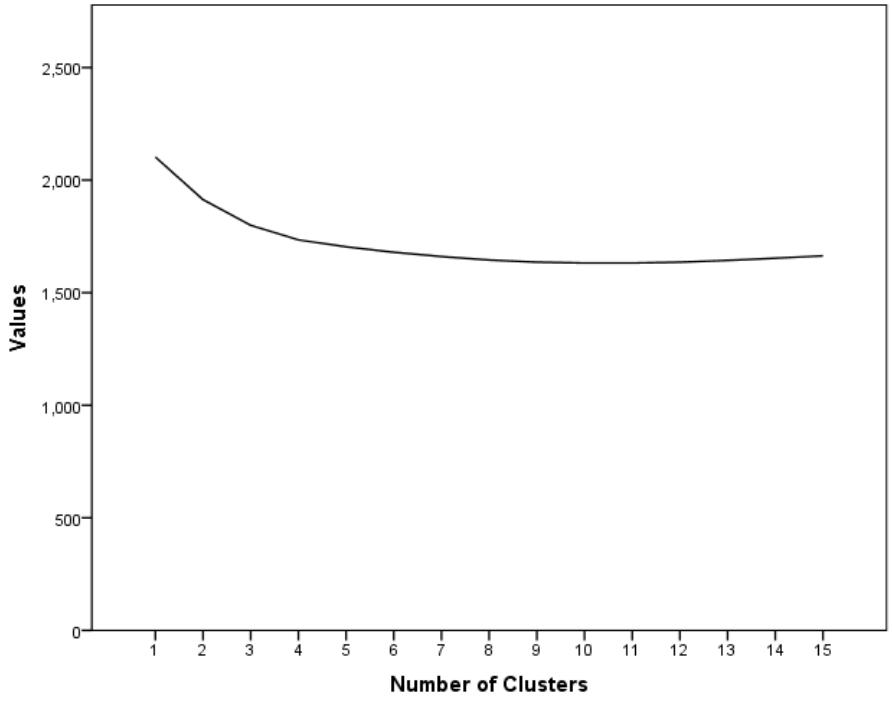


Figure 5-16 Line graph of Schwarz's Bayesian Criterion for RSI Two step cluster analysis

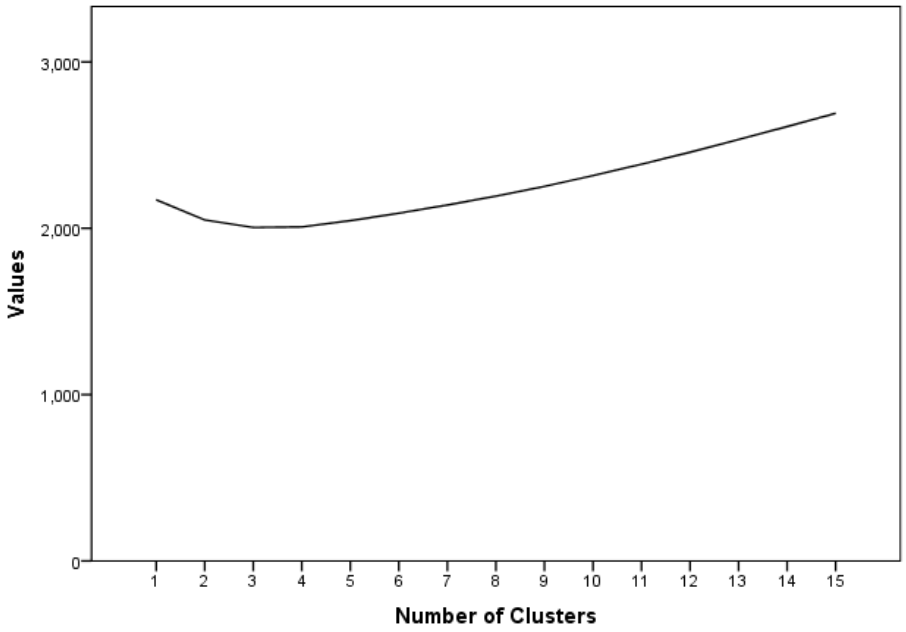
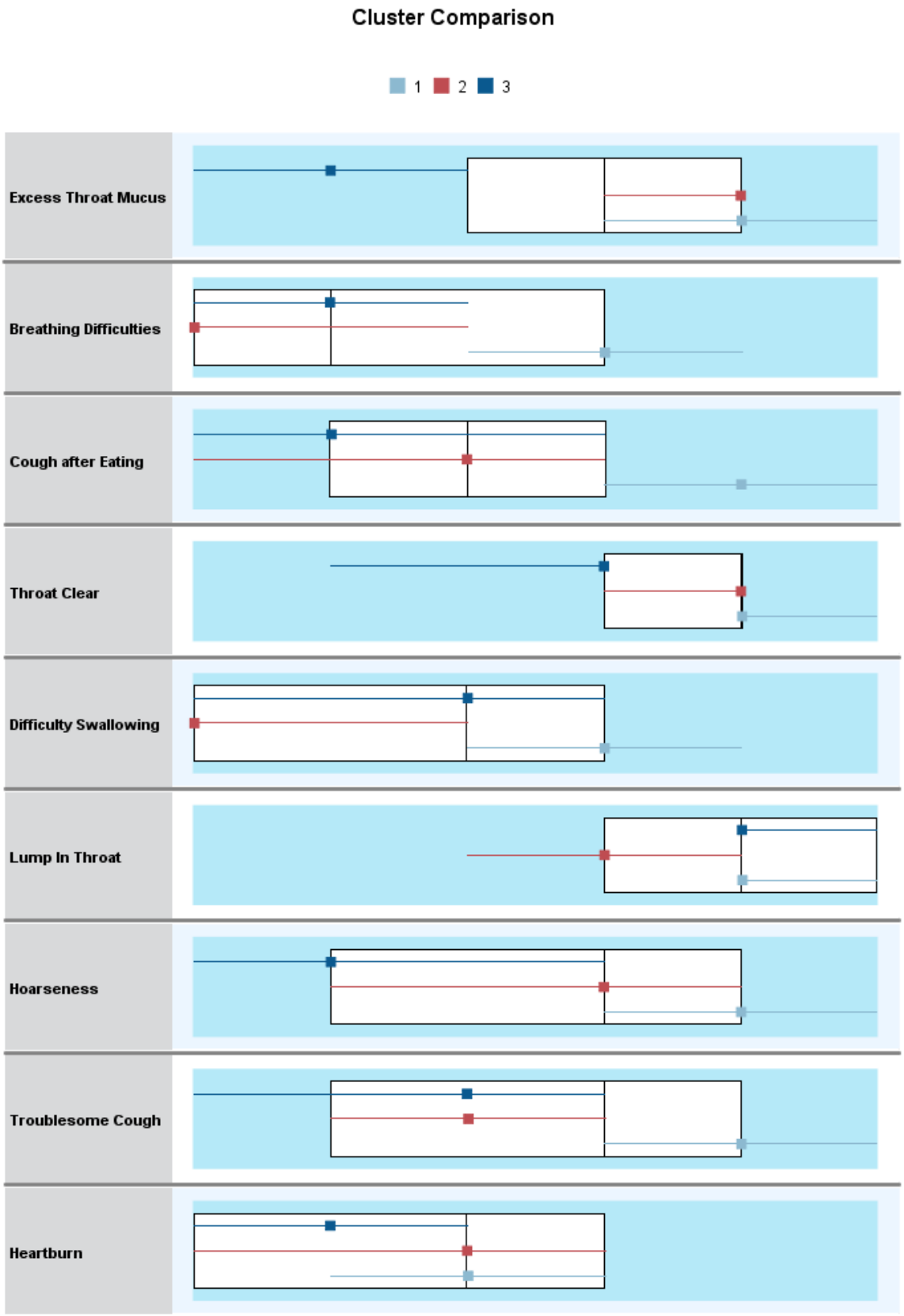


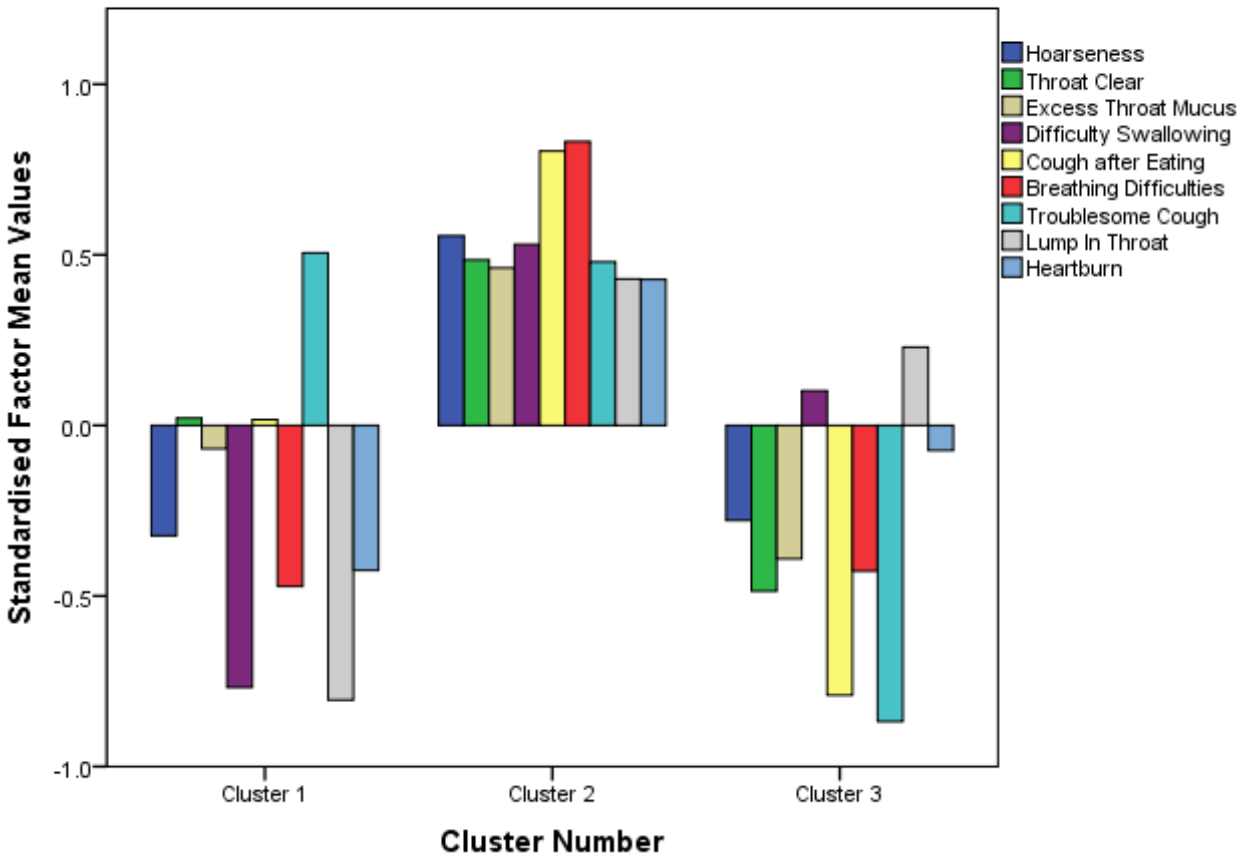
Figure 5-17 Two-step cluster analysis for the RSI



5.3.6.2 K-means RSI Cluster Analysis

Repeating the above methodology, standardised values for the nine RSI items were analysed. A two cluster model gave a simple split into high (n=194) and low (n=138) scoring clusters across all nine items when referenced with the TOPPITS population mean. A 3 cluster model gave a cluster of 96 patients scoring below the mean for all items other than cough, a cluster of 116 scoring above the mean for all items, and a cluster of 120 scoring below the mean other than for difficulty swallowing and something caught in the throat (Figure 5-18). The latter model converged within eight iterations and the ANOVA table showed that all nine items contributed significantly to the model.

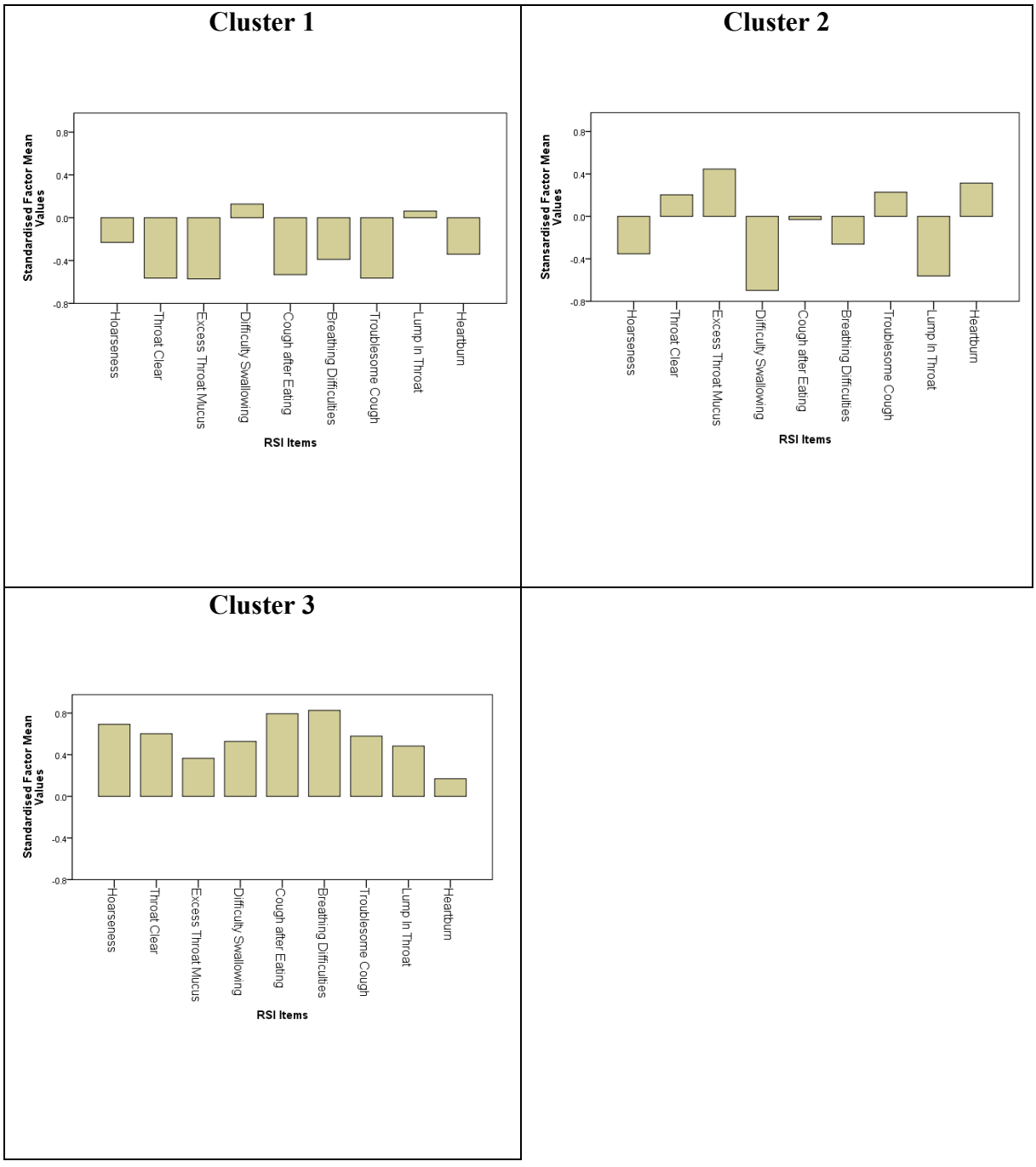
Figure 5-18 K-means three cluster model for RSI



5.3.6.3 Hierarchical cluster analysis of the RSI

With the number of clusters left unspecified, the analysis produced a two cluster model which was inappropriate, as described earlier. A three cluster model was therefore specified, in line with the previous results. This gave three clusters of 138, 98 and 96 patients. The raw data is presented in the appendix (Table 7-9,Table 7-10 and Table 7-11)and graphical representations below (Figure 5-19).

Figure 5-19 Hierarchical cluster analysis of the RSI



5.4 Discussion

For all solutions that either yielded or were specified to give a two cluster model, the two clusters were consistently split into high and low scoring clusters when compared to the overall population median or mean. This is not a particularly useful finding and is a reflection of the cluster analysis techniques splitting patients into low and high symptoms reporters, based around median scores. When all seven factors, or symptom groups, that emerged from the exploratory factor analysis in Chapter 4, were included in the cluster analysis, no clear pattern or clinically useful clustering solutions were produced. The analyses using the three highest factors appeared to offer clinically meaningful results.

The two step cluster analysis using the three highest factors with three clusters specified gave approximately equal sized clusters, and produced similar results to the K-means in suggesting that GI symptoms and voice reporting may offer the potential to distinguish patients' presenting symptoms. The two step analysis showed one cluster reporting higher cough and voice symptoms than the GI symptoms and one cluster reporting higher GI symptoms, when compared to the major cluster reporting low symptom scores across the three symptom groups. The cluster sizes were not the same between the K-means and two-step techniques, implying that not all patients mapped to the same cluster between the two methodologies. The advantage of the two-step analysis is the data spread that is depicted. In cluster one, the GI symptoms mean fell below the upper quartile of the population mean, whereas the cough and voice means were above the population means. However the range between the lower and upper quartiles for cluster one overlap amongst the three symptoms, which suggests that the GI symptoms were not significantly lower than the cough and voice symptoms. In cluster three the spread of the GI symptoms did appear to be higher than the spread for the cough and voice symptoms, which may suggest that for this cluster of 95 patients GI symptoms were reported significantly higher. This is the most that can be inferred when quartile ranges are used to compared the spread of data.

The use of voice symptom reporting as a potential distinguishing outcome emerged in the K-means analysis using the four factors with the highest mean factor scores. However, the most clinically useful solutions appeared when the three highest mean scored factors from the EFA were used; Voice, Cough and Gastrointestinal symptoms. The K-means analysis showed a cluster of patients reporting high gastrointestinal symptoms and a cluster reporting high voice symptoms, when compared to the major cluster with low scores across the three symptoms groups. One disadvantage of the K-means outputs is the inability to assess the spread of the data which means that limited conclusions can be drawn from bar charts which lack error

bars. The particular three factor K-means model of interest did not converge within 10 iterations, implying the analysis may not be appropriate to draw conclusions from.

The hierarchical cluster analysis of the Voice, Cough and GI symptoms, with three clusters specified, appeared to support the K-means and Two-step analyses. With the 95% confidence intervals manually calculated, and non-overlapping scores observed, this analysis gave a large cluster with lower scores in the three factors than the population mean, a moderate sized cluster with markedly higher GI symptoms, and a smaller cluster reporting higher voice symptoms. The K-means and hierarchical analyses therefore gave very similar cluster models, with the two-step analysis producing a model approaching this too.

Given the different presentations of the results from the three cluster analysis techniques a uniform graphical representation was required to assess if meaningful clusters were consistent across the techniques. The box and whisker graphs show that whilst voice and GI symptoms emerged as potential distinguishing symptoms between clusters, this finding was not apparent when the three techniques were assessed with one another.

The cluster analyses of the RSI questionnaire did not offer any clinically meaningful results. This is in keeping with the study by Papakonstantinou et al who found no identifiable clusters within the RSI and CReSS[44].

5.5 Conclusions

Cluster analysis techniques have not shown that clinically meaningful groups of patients can be defined from the TOPPITS baseline data.

Chapter 6 Thesis Discussion

This thesis has explored the wealth of presenting symptoms within a population of patients recruited to a large randomised clinical trial. The aims and objectives of the thesis were directed at only the baseline data from this trial. During the writing of this thesis, the final analysis of the TOPPITS trial was completed. No evidence was found to support the claim that stomach acid reduction medication is more effective than placebo in treating persistent throat and voice symptoms. Stomach acid reduction medications, in the form of proton-pump inhibitors (PPIs), are by far the commonest treatment for these patients. This thesis work has therefore taken on greater importance given the lack of evidence to support the efficacy of the current most popular method to manage these symptoms. Clinicians will need to readdress how to manage patients with persistent throat and voice symptoms, and with that will come a need to consider the symptoms in greater detail. This thesis materially adds to the evidence and understanding of these symptoms.

6.1 Statement of principal findings

Whilst not an aim of this thesis, chapter 2 demonstrates that the results from the TOPPITS trial can be applied to any general population of patients with persistent throat and voice symptoms. The patients' demographics and outcome measures scores from the TOPPITS baseline data have been described and are comparable to the published literature. The highest scored symptoms were those of a lump in the throat, throat clearing and excess throat mucus. Traditional heartburn symptoms were not ranked highly and this finding is again in keeping with the published literature. The baseline TOPPITS data would therefore appear generalisable. As such, the findings from this thesis could also be applied to any other population of patients with persistent throat symptoms.

The RSI, CReSS and LPR-HRQL are related in terms of total scores from this patient population. In itself this is not surprising, given the overlap and repetition in the symptoms that they cover. It is highly unlikely that a clinical trial will ever use these questionnaires in this combination again. At the time of the TOPPITS trial design, the RSI appeared the most popular questionnaire to assess throat symptoms. However, clinicians had concerns over its brevity, the polysymptomatic reflux question and the fact that previous studies of proton-pump inhibitors with positive outcomes may have reported such results due to improvements in these reflux symptoms alone[56]. The CReSS was used in the trial to include more detailed symptoms in its 34 items. The LPR-HRQL was included as it was the only patient reported outcome measure specifically aimed at disease-specific quality of life.

The demographic data analysis also demonstrated that increasing age was associated with reduced symptom score reporting. This was found with the CReSS and LPR-HRQL but not with the RSI. This appears to be a consistent finding with similar results reported in previous studies using the RSI [27, 60].

There was a significant number of patients with missing RFS scores in the baseline data analysis. This was discussed by the TOPPITS' trial management group, who agreed this was not due to reporter bias, but rather the throat images were missing for groups of consecutive patients from particular institutions. Nevertheless, the TOPPITS RFS data offers by far the largest analysis of throat and laryngeal images. In contrast to the most popular cited study of 40 patients, which showed high concordance between symptoms and throat signs[26], this thesis found no association between any of the three questionnaires and the RFS scores in 254 patients. The RFS was scored by an experienced academic speech and language therapist who was blind to the questionnaire scores. Some concerns may be raised as the equipment used to capture the images of the throat was not standardised across recruiting centres. The lack of a relationship between patient reported symptoms and clinician rated throat appearances calls into question the popularity of laryngopharyngeal reflux as a diagnosis.

Schneider at al [30] assessed the evidence of a causal relationship between reflux and voice changes. They used the Bradford-Hill criteria and concluded that the evidence was insufficient to prove causality. As described in the introduction to this thesis there is no definition for a diagnosis of LPR. Intuitively one would assume the diagnosis would be made using a combination of patient reported symptoms, clinical examination and specialist investigations. Objective measures of gastroesophageal reflux include pH manometry and impedance testing. These measure the pH, frequency and timings of reflux episodes from the stomach into the oesophagus. Access to these investigations is often limited and in the UK are more widely adopted by surgeons considering gastric anti-reflux surgery. They are not in common use in the UK to assess patients with throat and voice symptoms. There is a lack of dose –response and temporal relationships between these reflux investigations and symptoms [30]. Objective measures of reflux were not used in the TOPPITS trial so as to preserve the pragmatic nature of the clinical trial and to maximise recruitment to the study. This thesis found no evidence of a relationship between clinical examination findings and patient reported symptoms. In essence this suggests that the clinical examination of these patients should be focused on excluding defined pathological causes for symptoms, and should not focus on throat and laryngeal signs that have been attributed to reflux – as set out in the RFS. It would be wise to treat patients on the merits of their symptoms alone, and not on their

laryngeal and throat appearances. That is to say, incidental laryngeal appearances found in patients who lack persistent throat symptoms should not result in a diagnosis of reflux. The lack of coherence between symptoms, signs and investigations should mean that reported symptoms form the basis of the clinical consultation and management strategy.

Gastroesophageal reflux is associated with increasing BMI [61]. This thesis found no relationship between throat and voice symptom reporting and BMI. It did find that increasing BMI might be related to higher RFS scores. However, this was a weak relationship and the lack of coherence otherwise between symptoms, signs and BMI means drawing conclusions from this one finding would be inappropriate. The lack of any relationship between BMI and throat symptoms in this thesis may again lead us to question whether it is correct to theorise that reflux causes throat and voice symptoms.

Exploratory factor analysis demonstrated seven factors of variables (symptoms) using the combined data from the RSI, CReSS and LPR-HRQL. This offers a potential clinically meaningful and simplified classification of persistent throat symptoms: voice, cough, gastrointestinal symptoms, airway symptoms and dysphagia, throat clearing, life events, and lump in the throat sensation The dimension reduction produced through this methodology could be useful in future research, such as defining an optimal patient reported outcome tool. Exploratory Factor Analysis of the individual questionnaires gave poor models for the RSI, which was not in agreement with previously published data. The factor structure for the CReSS was clean and reproduced the Oesophageal, Upper Airway and Pharyngeal factors previously described for this questionnaire[45]. The LPR-HRQL main symptom domains do contain one factor each, other than the throat domain, which suffers with an item covering both voice and throat clearing, and three questions covering life events which consistently load together as a separate factor.

Cluster analysis demonstrated separate groups of patients within the TOPPITS population that reported a higher burden of gastrointestinal symptoms or a higher burden of voice symptoms. However, when the three techniques were compared this finding was not consistent. No clinically meaningful clusters of patients were defined. It is therefore reasonable to conclude that individual throat and voice symptoms cannot be used to categorise groups of patients reliably. A global term that encompasses the range of interlinked symptoms, such as persistent throat symptoms, would appear more appropriate than referring to individual symptoms. This thesis and the TOPPITS results imply that a broad term like persistent throat symptoms is more appropriate to recommend to clinicians than LPR.

This thesis has found no definable groups, based on symptoms, which could be used to justify subgroup analysis of the TOPPITS final outcomes.

6.2 Strengths and weaknesses of the study

The TOPPITS baseline data represents by far the largest sample of patients with persistent throat symptoms for whom rigorous prospective outcomes have been collected. The volume of missing data was very low. The large sample size allows robust conclusions to be drawn from the analyses within this thesis. Concentrating the thesis on only the baseline data within TOPPITS has ensured that the presenting symptoms have been analysed independent to any of the TOPPITS results.

Patients entering clinical trials may not represent the general population. We had some concerns that patients who had used proton-pump inhibitors previously for heartburn complaints may not enter the trial. The ninth item of the RSI, covering traditional reflux symptoms, was removed from the inclusion criteria so patients needed to score above 10 on the non-heartburn items of the questionnaire. This was done as previous reports of positive trial outcomes with PPIs, attributed these results to improvements in the heartburn symptoms alone on the RSI[56]. The demographic analysis in chapter 2 clarifies that there is no major discrepancy between the TOPPITS population and other populations of persistent throat symptoms patients, as evidenced through all-inclusive large case series data, in regards symptom reporting and ranking of RSI items. Whilst the exclusion of patients with strong heartburn symptoms was a concern, this does not appear to have been born out. As such, the thesis results appear applicable to other populations of patients with persistent throat symptoms.

The Exploratory Factor and Cluster analyses have proven useful methods to group symptom variables and patients. However, the results from these were influenced by the questionnaire constructs. In many ways, the seven factors from the EFA could be viewed as resembling the RSI items. The voice factor had the highest mean loading score on the EFA and the greatest number of items within a single factor. That was influenced by the 12 voice items from the LPR-HRQL. In contrast, the RSI and CReSS only had one voice item each. If the LPR-HRQL questionnaire had not been included, then the voice factor may not exist as a single entity but may rather load with the other upper airway items. This would have altered the subsequent cluster analysis.

The use of three cluster analysis techniques is not common but can be considered a strength of the study. Viewed as independent analyses, the thesis would have concluded that voice and GI

symptoms may define groups of patients for subgroup analyses in the TOPPITS trial. However when viewed alongside each other the techniques' results clearly do not offer a consistent finding.

6.3 Unanswered questions and future research

This thesis has explored presenting throat and voice symptoms. It has not assessed the most suitable patient reported outcome measure for this set of symptoms or condition. There are clear methods for assessing the utility of PROMs[46]. Now that the TOPPITS trial has completed, a future study could reassess the three PROMs used and measure, amongst other criteria, the response to change over time or the inter-rater reliability. From the work performed in this thesis, the CReSS questionnaire had a more definable factor structure than the RSI, which was also very consistent with a previous EFA[45]. However, a number of items were omitted for low loading scores from the EFA, and given the thesis results, perhaps more emphasis on voice symptoms should be considered. Only one of the 34 items relates to hoarseness. There were no major concerns with the LPR-HRQL, but some items cover more than one symptom, and only three items were separated as a potential "life events" factor, with all the others loading with the symptoms as set out in the RSI or CReSS. So perhaps fewer items pertain to disease specific quality of life than were intended. Any tool used should be easily managed by patients and clinicians. The structure of the LPR-HRQL and scoring system appears unwieldy. A follow on project could assess the most suitable PROM for future studies to adopt based on the checklist of criteria and the performances of these questionnaire within this clinical trial. These results could then form the basis to commence a Core Outcomes Set initiative for patients with persistent throat symptoms.

A further research project could assess the questionnaire constructs in greater detail, incorporating patients' views. Patient involvement in the questionnaire designs has been lacking, and whilst we can hypothesise about for instance how many voice items should be included in a 34 item questionnaire, the best way to address this would be through the thoughts of the patients. There is, understandably, great variation and subjectivity in how patients view the terminology and questions relating to persistent throat symptoms. Medical terminology has changed from "unexplained symptoms" to "persistent symptoms" on the basis of patients' input. As the introduction chapter explained, the semantics and interpretation over symptoms such as coughing, throat clearing and post nasal secretions are not clear. Clarifying patients' interpretation over these symptoms would be an important first step in this research.

The thesis results can now be taken in context with the TOPPITS results. Many clinicians have asked the TOPPITS trial team how they should now manage patients with persistent throat symptoms. In response to an NIHR call for Policy Relevant Research, I led a multidisciplinary application in October 2019 entitled "Redefining the Management of Persistent Throat Symptoms". The research plan was to disseminate the TOPPITS results within primary and specialist secondary care and attempt to offer an alternative patient-centred approach to managing throat symptoms. Much of the basis for this approach would be guided by results from this thesis. We recognise the challenges in altering prescribing practices of clinicians, and that a simple message of not giving PPIs will not be as strong as a message that offers an alternative strategy. Unfortunately, this bid was unsuccessful, but the question of "what next" for these patients in the post-PPI era is clearly of great importance.

Finally, the results of this thesis are directly relevant to the TOPPITS team. At the time of writing this thesis discussion, the emphasis is on publishing the main trial paper and the Health Technology Assessment report for the trial. TOPPITS was not designed to perform subgroup analysis. The results found no evidence in favour of a PPI over placebo. It may therefore be inappropriate to consider any further exploratory subgroup analyses of the trial outcomes. This thesis has found no definable clusters of patients to justify subgroup analysis.

As this thesis is finalised in January 2021, the TOPPITS results are due to be published in the British Medical Journal (BMJ). The journal editors have requested that we write an article for BMJ Opinions on the topic of persistent throat symptoms and their management. This is very timely as the BMJ recently published an educational article aimed at primary care, which again recommended PPIs for throat symptoms such as globus sensation. This article generated some useful social media discussions. During the current Covid-19 pandemic, doctors have consulted patients remotely and we have seen an anecdotal rise in the use of PPIs for throat symptoms in primary care. The BMJ is the ideal platform to highlight the TOPPITS results to primary care clinicians, and then to allow us to offer an alternative management strategy for this large group of patients. Recommending a broad term which removes reflux as a causative element from the title, such as "persistent throat symptoms", will be an important angle to cover in changing practice away from PPIs. Persistent throat symptoms could be defined as six weeks or more of symptoms that include: globus sensation, intermittent dysphonia, throat clearing, catarrh (or "post-nasal drip"), cough (from the larynx level or throat – not lungs), and throat discomfort.

We acknowledge that further research is required to clarify the role of other anti-reflux treatments in managing persistent throat symptoms, such as alginates or life style advice; which may include weight loss, avoidance of eating late before bed and raising the head of the bed. Offering doctors strategies that involve proven patient delivered therapies to reduce throat irritation, throat dryness and throat clearing will be a marked shift from PPI prescriptions and will require promotion through a variety of media and treatment guidelines.

Appendices

Research Ethics Committee Approval

NHS
Health Research Authority

East of England - Essex Research Ethics Committee

The Old Chapel
Royal Standard Place
Nottingham
NG1 6FS

Please note: This is the favourable opinion of the REC only and does not allow you to start your study at NHS sites in England until you receive HRA Approval

14 May 2018

Mr James O'Hara

Consultant Otolaryngologist

Newcastle upon Tyne Hospitals NHS Foundation Trust

Department of Ear Nose and Throat

The Freeman Hospital

Newcastle upon Tyne

NE7 7DN

Dear Mr O'Hara

Study title:	**Characterisation of patients presenting with persistent throat symptoms**
REC reference:	**18/EE/0158**
Protocol number:	**NCTU:6831**
IRAS project ID:	**240544**

The Proportionate Review Sub-committee of the East of England - Essex Research Ethics Committee reviewed the above application on 08 May 2018.

We plan to publish your research summary wording for the above study on the HRA website, together with your contact details. Publication will be no earlier than three months from the date of this favourable opinion letter. The expectation is that this information will be published for all studies that receive an ethical opinion but should you wish to provide a substitute contact point, wish to make a request to defer, or require further information, please contact hra.studyregistration@nhs.net outlining the reasons for your request. Under very limited circumstances (e.g. for student research which has received an unfavourable opinion), it may be possible to grant an exemption to the publication of the study.

Ethical opinion
On behalf of the Committee, the sub-committee gave a favourable ethical opinion of the above research on the basis described in the application form, protocol and supporting documentation, subject to the conditions specified below.

Conditions of the favourable opinion

The REC favourable opinion is subject to the following conditions being met prior to the start of the study.

Committee comment

> The Committee commented that the document submitted as the Protocol was sufficient in this instance but wished to remind the applicant that, for any future research, a formal protocol would be required

Management permission must be obtained from each host organisation prior to the start of the study at the site concerned.

Management permission should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements. Each NHS organisation must confirm through the signing of agreements and/or other documents that it has given permission for the research to proceed (except where explicitly specified otherwise).

Guidance on applying for HRA and HCRW Approval (England and Wales)/ NHS permission for research is available in the Integrated Research Application System, www.hra.nhs.uk or at http://www.rdforum.nhs.uk.

Where a NHS organisation's role in the study is limited to identifying and referring potential participants to research sites ("participant identification centre"), guidance should be sought from the R&D office on the information it requires to give permission for this activity.

For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.

Sponsors are not required to notify the Committee of management permissions from host organisations.

To ensure transparency in research, we strongly recommend that all research is registered but for non-clinical trials this is not currently mandatory.

If a sponsor wishes to request a deferral for study registration within the required timeframe, they should contact hra.studyregistration@nhs.net. The expectation is that all clinical trials will be registered, however, in exceptional circumstances non registration may be permissible with prior agreement from the HRA. Guidance on where to register is provided on the HRA website.

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

Ethical review of research sites

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see
"Conditions of the favourable opinion").

Summary of discussion at the meeting

The PR Sub-Committee agreed that this was a well presented study with no material ethical issues.

Committee comment

The Committee commented that the document submitted as the Protocol was sufficient in this instance but wished to remind the applicant that, for any future research, a formal protocol would be required

Approved documents

The documents reviewed and approved were:

Document	*Version*	*Date*
IRAS Application Form [IRAS_Form_19042018]		19 April 2018
Letter from sponsor [Sponsor Letter]		09 April 2018
Other [Wilson CV]		28 March 2018
Other [Fisher CV]		29 March 2018
Research protocol or project proposal [project proposal 19_3_2018]		19 March 2018
Summary CV for Chief Investigator (CI) [CI CV]		02 March 2018
Summary CV for supervisor (student research) [supervisor's CV]		28 March 2018

Membership of the Proportionate Review Sub-Committee

The members of the Sub-Committee who took part in the review are listed on the attached sheet.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

Reporting requirements

The attached document "After ethical review – guidance for researchers" gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Notification of serious breaches of the protocol
- Progress and safety reports

- Notifying the end of the study

The HRA website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

User Feedback

The Health Research Authority is continually striving to provide a high quality service to all applicants and sponsors. You are invited to give your view of the service you have received and the application procedure. If you wish to make your views known please use the feedback form available on the HRA website:
http://www.hra.nhs.uk/about-the-hra/governance/quality-assurance/

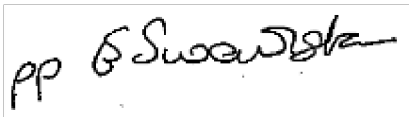
HRA Training

We are pleased to welcome researchers and R&D staff at our training days – see details at http://www.hra.nhs.uk/hra-training/

With the Committee's best wishes for the success of this project.

18/EE/0158	**Please quote this number on all correspondence**

Yours sincerely



Dr Gerry Kamstra
Vice Chair

Email: NRESCommittee.EastofEngland-Essex@nhs.net

Enclosures: *List of names and professions of members who took part in the review*

"After ethical review – guidance for researchers"

Copy to: *Mrs Kay Howes*

East of England - Essex Research Ethics Committee

The Outcome Measures

The Reflux Symptom Index

Within the last MONTH, how did the following problems affect you?

0 = no problem 5 = severe problem

Hoarseness or a problem with your voice	0	1	2	3	4	5
Clearing your throat	0	1	2	3	4	5
Excess throat mucus or postnasal drip	0	1	2	3	4	5
Difficulty swallowing food, liquids or tablets	0	1	2	3	4	5
Coughing after eating or lying down	0	1	2	3	4	5
Breathing difficulties or choking episodes	0	1	2	3	4	5
Troublesome or annoying cough	0	1	2	3	4	5
Sensation of something caught in your throat or a lump in your throat	0	1	2	3	4	5
Heartburn, indigestion or stomach acid coming up	0	1	2	3	4	5

TOTAL		**TOTAL minus heartburn score**	

The Comprehensive Reflux Finding Score

Thank you for agreeing to be in this study. Please complete the following questions.

Within the last month, how did the following problems affect you?

Please tick the box that best fits.

Symptom	***0 No problem***	***1***	***2***	***3***	***4***	***5 Severe problem***
heartburn						
pressure in chest						
regurgitation						
acid/sour taste in mouth						
gurgling stomach						
pressure/lump in throat						
difficulty swallowing food						
difficulty swallowing liquids						
nausea						
pain in throat						
vomiting						
bloating						
belching						
flatulence						
hiccups						
decreased appetite						
rush of saliva into mouth						
feeling full too early in a meal						
bad breath						
back pain						
headache						
choking						

Symptom	*0 No problem*	*1*	*2*	*3*	*4*	*5 Severe problem*
coughing when upright						
coughing after eating						
coughing when lying down						
wheezing						
difficulty breathing						
hoarseness						
throat clearing						
excess mucus						
mucus dripping down back of throat						
feeling things stuck in throat						
indigestion						
stomach acid coming up						
other – please describe						

The Laryngopharyngeal Reflux Health Related Quality of Life Questionnaire

LPR-QOL Questionnaire –©AstraZeneca LP 2003

Here are some questions about laryngopharyngeal reflux (L R, acid reflux into your upper throat, silent reflux) and how it affects you.

Below most questions you will see numbers from 0 to 6, with a description of how often you experience that symptom. If you have a symptom that is described in a question, even if you think it is caused by something else (for example, allergies or a cold), please go ahead and answer the question anyway.

If a symptom does not apply to you, then you would circle "0," "None of the time". On the other and, if you experienced a symptom all of the time or always, you would circle "6," "Nearly all of the time or always."

You will notice that after a set of questions that relate to a particular symptom, there is a question in bold type rated on a scale from 1 to 10 which asks you to summarize the overall impact of those symptoms on your life.

Finally, at the end of the questionnaire, there are a number of questions that use the 1 to 10 scale. Please be sure to complete each question by circling the number and description that best states your answer. Often your first response that comes to mind is the best one.

There are no right or wrong answers, and that your answers to the questionnaire will be kept confidential.

Don't forget to answer every question!

Thank you

For the purposes of editing, the scoring descriptions have been removed from the table, as they would normally appear, and shown below.

Score or LPR-HRQL	**Description of score in questionnaire**
0	None of the time (never in the past month)
1	Rarely (once in the past month)
2	A little of the time (2-3 days in the past month)
3	Some of the time (about once a week)
4	A lot of the time (about 2 to 3days a week)
5	Most of the time (4-5 days a week)
6	Nearly of the time or always (6-7 days a week)

OVER THE PAST 4 WEEKS…..

Please tick the appropriate box.

	0	**1**	**2**	**3**	**4**	**5**	**6**
My Voice Problems Make It Difficult For Me To Work							

I Feel Satisfied With The Way My Voice Sounds							
Being Hoarse Makes It Hard For Me To Communicate My True Self							
My Voice Makes Others Feel Uncomfortable To Listen To Me when I am talking							
I Can't Sing As Much As I Would Like To Because Of My Voice							
I Find It Hard To Meet New People Because Of What They Will Think							
The Sound Of My Voice Makes People Think I'm Angry Or Upset							
Straining To Talk Is Tiring							
I Feel Embarrassed About The Sound Of My Voice							
I Avoid Talking Because Of The Effort							
My Voice Problems Make It Difficult For Me To Do My Job							
I Am Afraid I Might Lose My Voice Forever							

Questions 1-12 above were about talking, singing, and your voice. The question below asks you to think about how problems with your voice affected your overall quality of life.

Please rate from 1 to 10, where 1 means "no effect" and 10 means 'an enormous effect," how much of an effect have problems with talking, singing and your voice had on your overall quality of life? *(circle one).*

1 2 3 4 5 6 7 8 9 10

OVER THE PAST 4 WEEKS…..

Please tick the appropriate box.

	0	**1**	**2**	**3**	**4**	**5**	**6**

My Coughing Embarrasses Me							
I Avoid Social events Where My Coughing Might Bother Others							
I Have To Leave The Room Because Of My Coughing							
People Think I Am Sick Because Of My Coughing							
My Co-workers Can Hear Me Coming Because Of My Coughing							
I Worry About Having A Coughing Spell At A Bad Time							

Questions 14-19 were about coughing. The question below asks you to think about how problems with coughing affected your overall quality of life.

Please rate from 1 to 10, where 1 means "no effect" and 10 means an "enormous effect," how much of an effect have coughing had on your overall quality of life? *(circle one).*

1 2 3 4 5 6 7 8 9 10

OVER THE PAST 4 WEEKS…..Please tick the appropriate box.

	0	**1**	**2**	**3**	**4**	**5**	**6**
People Notice How Much I Have To Clear My Throat							
Clearing My Throat Has A Negative Effect On Sex							
Clearing My Throat Has A Negative Effect On Friendships							
The Need To Clear My Throat Makes It Hard To Talk							

Questions 21-26 above were about clearing your throat. The question below asks you to think about how problems with clearing your throat affected your overall quality of life.

Please rate from 1 to 10, where 1 means "no effect" and 10 means an "enormous effect," how much of an effect problems with clearing your throat had on your overall quality of life? *(circle one).*

1 2 3 4 5 6 7 8 9 10

OVER THE PAST 4 WEEKS…..

Please tick the appropriate box.

	0	**1**	**2**	**3**	**4**	**5**	**6**
I Feel A Lump In My Throat Which Makes It Hard To Swallow							
I Hesitate To Eat In Public Because I Have Trouble Swallowing							
I Am Afraid Of Choking In My Sleep							
I Am Bothered By A Burning Feeling In My Throat							
I Awaken From Sleep Gasping For Breath							

Questions 28-32 above were about general throat symptoms. The question below asks you to think about how problems with general throat symptoms affected your overall quality of life.

Please rate from 1 to 10, where 1 means "no effect" and 10 means an "enormous effect," how much of an effect problems with your throat had on your overall quality of life? *(circle one).*

1 2 3 4 5 6 7 8 9 10

The remaining questions refer to the combined impact of all of your acid reflux related symptoms (acid reflux into your upper throat) on various aspects of your life.

34. In general, on a scale from 1 to 10, where 1 means "no effect" and 10 means "an enormous effect," how much of an effect have symptoms described above had on your energy levels? (*circle one)*

1 2 3 4 5 6 7 8 9 10

35. On a scale from 1 to 10, where 1 means "no effect" and 10 means "an enormous effect," how much of an effect have symptoms described above had on your productivity at your job/work? (*circle one)*

1 2 3 4 5 6 7 8 9 10

36. On a scale from 1 to 10, where 1 means "no effect" and 10 means "an enormous effect," how much of an effect have symptoms described above had on your social relationships? (*circle one)*

1 2 3 4 5 6 7 8 9 10

37. On a scale from 1 to 10, where 1 means "no effect" and 10 means "an enormous effect," how much of an effect have symptoms described above had on your marital/intimate relationships? (*circle one)*

1 2 3 4 5 6 7 8 9 10

38. On a scale from 1 to 10, where 1 means "no effect" and 10 means "an enormous effect," how much of an effect have symptoms described above had on your sexual relationships?

1 2 3 4 5 6 7 8 9 10

39. On a scale from 1 to 10, where 1 means "no effect" and 10 means "an enormous effect," how much of an effect have symptoms described above had on your sleeping? (*circle one)*

1 2 3 4 5 6 7 8 9 10

40. On a scale from 1 to 10, where 1 means "no effect" and 10 means "an enormous effect," how much of an effect have symptoms described above had on your ability to lie comfortably in bed? (*circle one)*

1 2 3 4 5 6 7 8 9 10

41. On a scale from 1 to 10, where 1 means "no effect" and 10 means "an enormous effect," how much of an effect have symptoms described above had on the way you feel about yourself? (*circle one)*

1 2 3 4 5 6 7 8 9 10

42. On a scale from 1 to 10, where 1 means "no effect" and 10 means "an enormous effect," how much of an effect have symptoms described above had on your lifestyle (i.e. smoking, drinking, exercise, eating habits)? (*circle one)*

1 2 3 4 5 6 7 8 9 10

43. On a scale from 1 to 10, where 1 means "no interference" and 10 means "a great deal of interference" how much of an effect have symptoms described above interfered with your ability to do the things you enjoy? (*circle one)*

1 2 3 4 5 6 7 8 9 10

Thank you for completing this questionnaire

The Reflux Finding Score

Subglottic Oedema	0 = absent 2 = present
Ventricular Obliteration	2 = partial 4 = complete
Erythema/hyperaemia	2 = arytenoids only 4 = diffuse
Vocal fold Oedema	1 = mild 2 = moderate 3 = severe 4 = polypoid
Diffuse laryngeal Oedema	1 = mild 2 = moderate 3 = severe 4 = obstructing
Posterior commissure hypertrophy	1 = mild 2 = moderate 3 = severe 4 = obstructing
Granuloma/granulation tissue	0 = absent 2 = present
Thick endolaryngeal mucus	0 = absent 2 = present

Demographic Data Analysis

Figure 6-1 Histogram and overlying normal curve showing the distribution of the RSI item 1 - Hoarseness

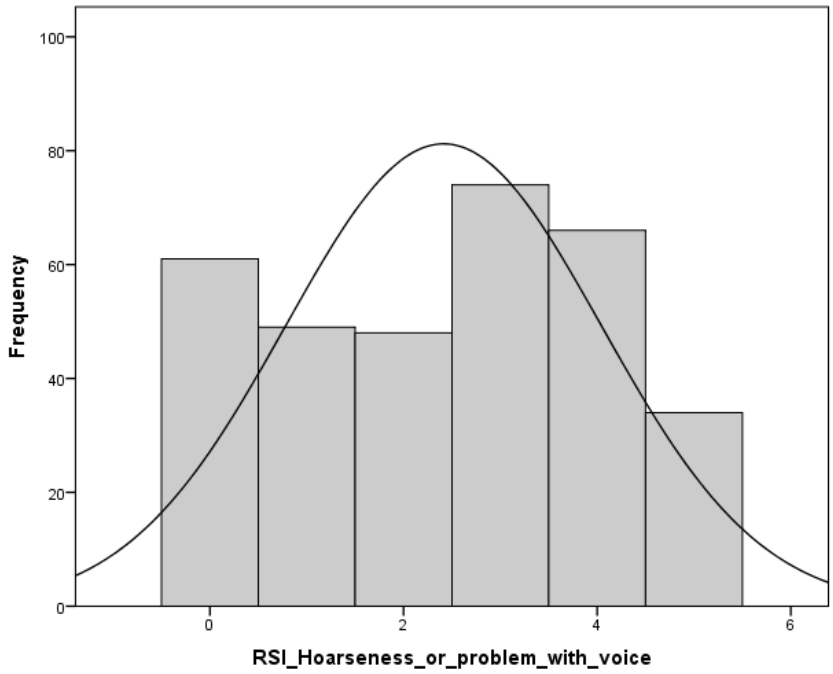


Figure 6-2 Histogram and overlying normal curve showing the distribution of the RSI item 2 - Throat clearing

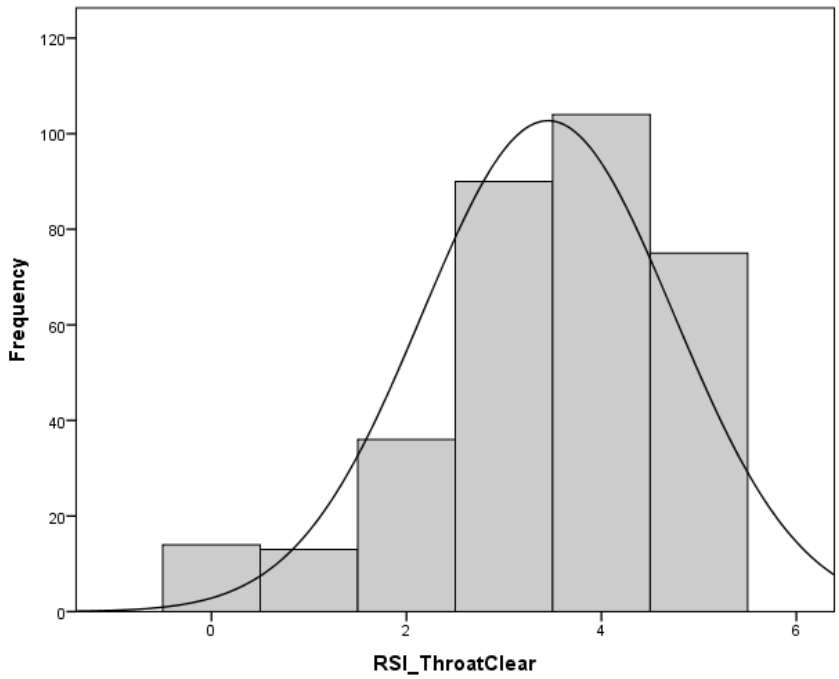


Figure 6-3 Histogram and overlying normal curve showing the distribution of the RSI item 3 - Difficulty swallowing

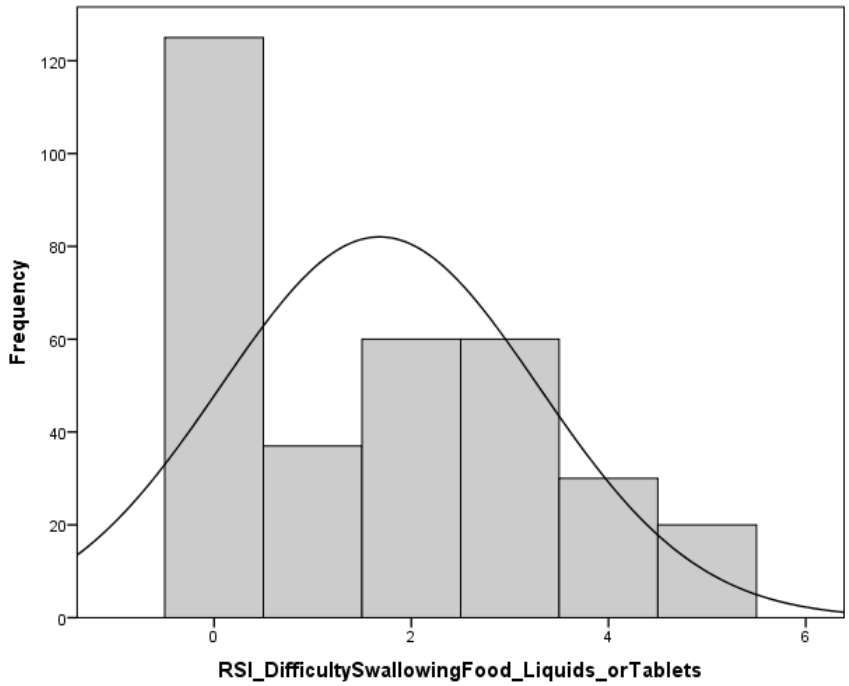


Figure 6-4 Histogram and overlying normal curve showing the distribution of the RSI item 4 - Excess Throat Mucus and Post Nasal Drip

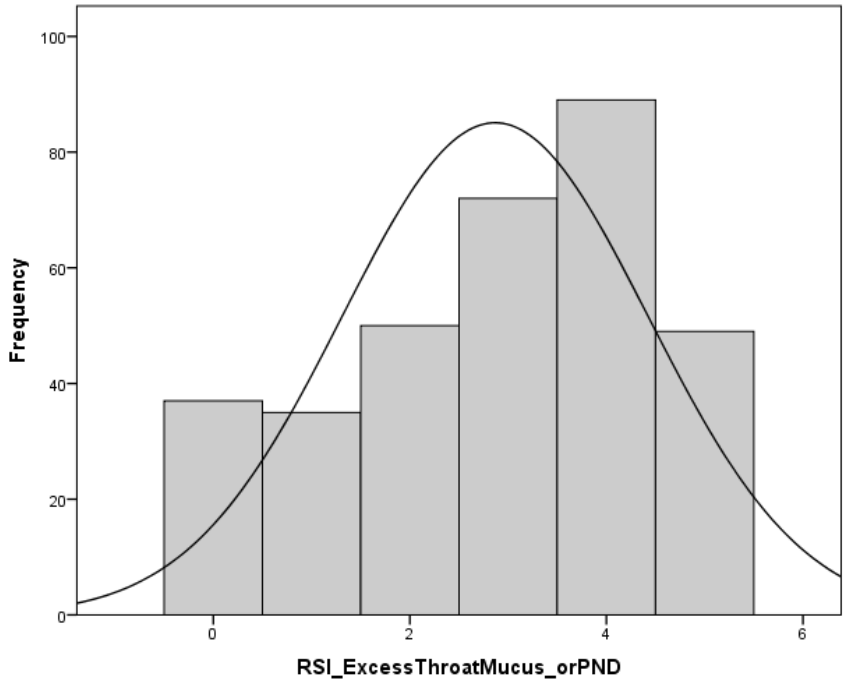


Figure 6-5 Histogram and overlying normal curve showing the distribution of the RSI item 5 - Coughing after eating / lying

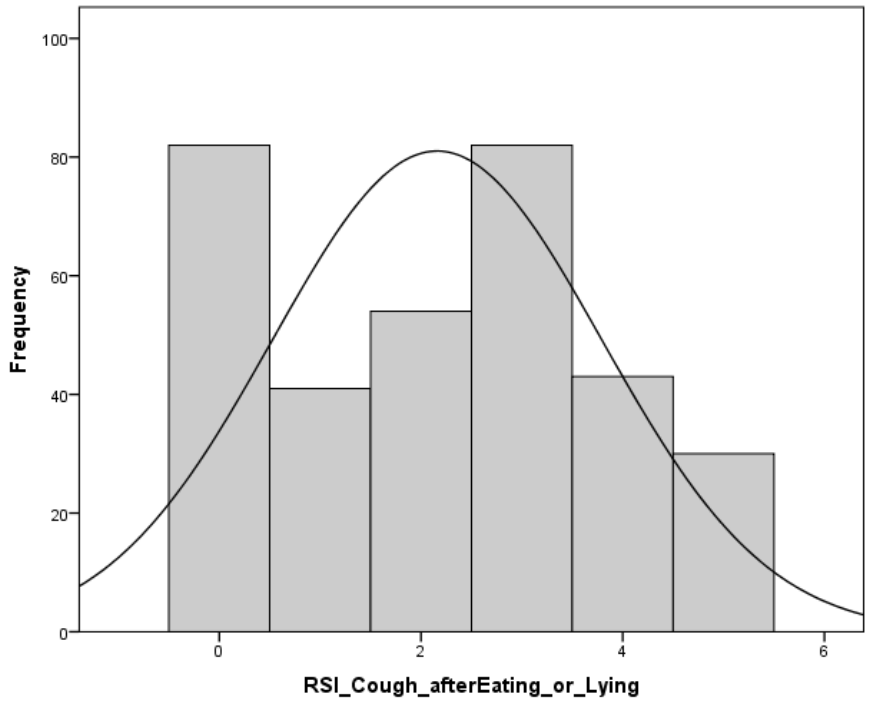


Figure 6-6 Histogram and overlying normal curve showing the distribution of the RSI item 6 - Breathing difficulties / choking

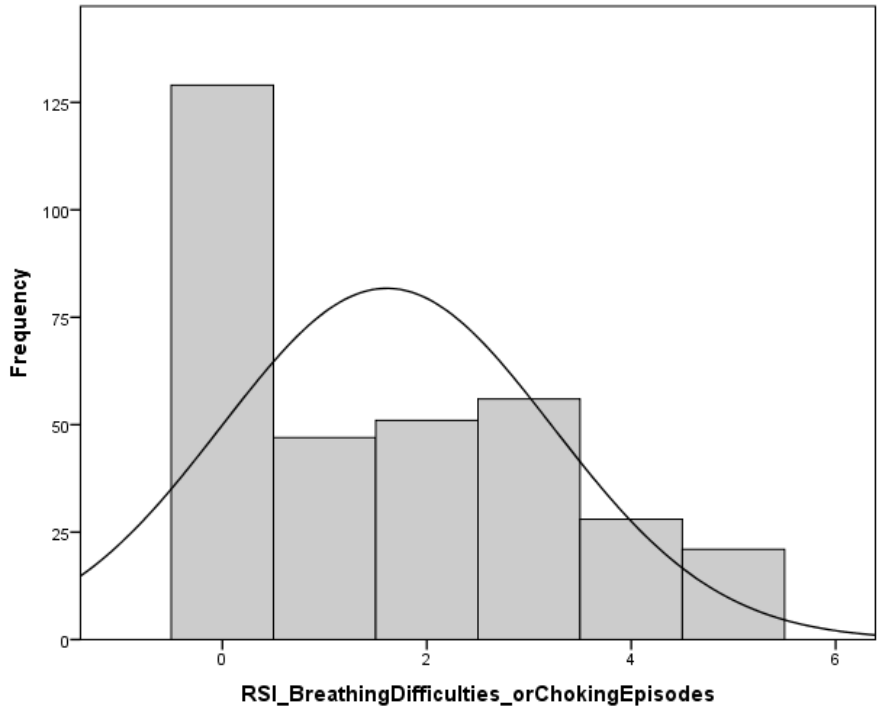


Figure 6-7 Histogram and overlying normal curve showing the distribution of the RSI item 7 - Troublesome cough

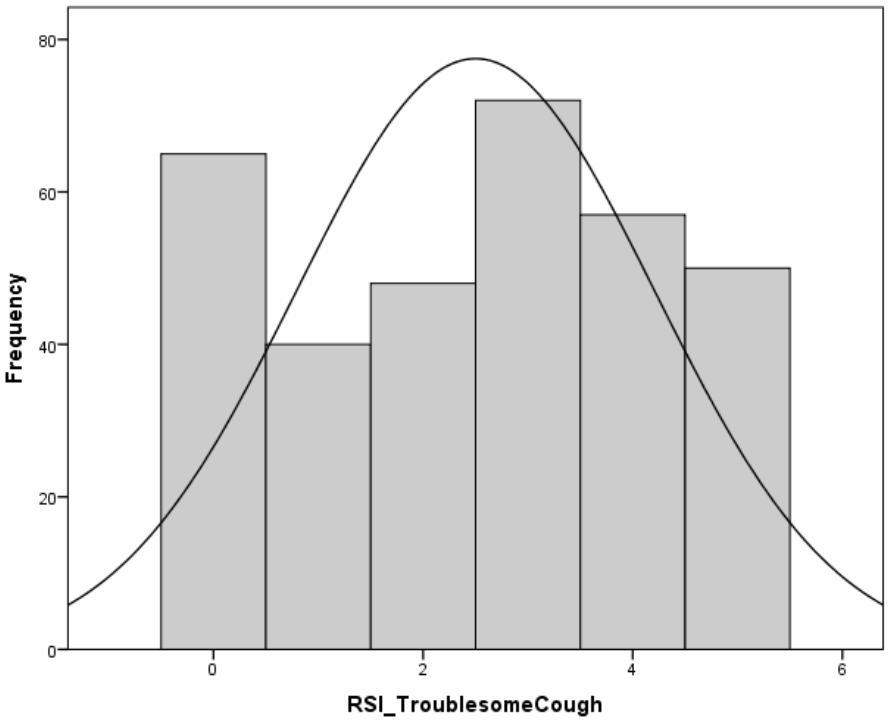


Figure 6-8 Histogram and overlying normal curve showing the distribution of the RSI item 8 - Lump in throat

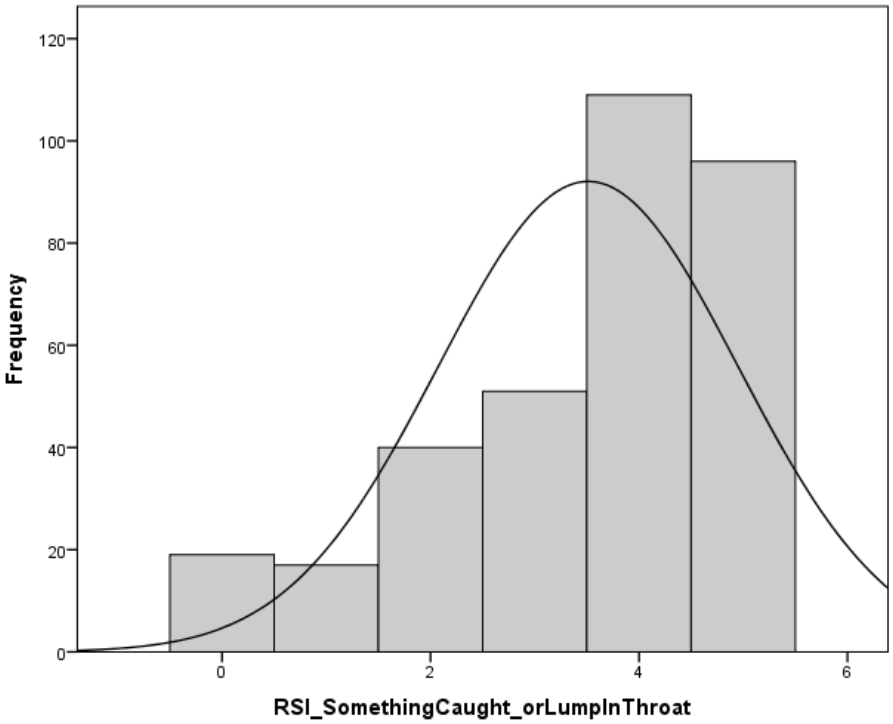
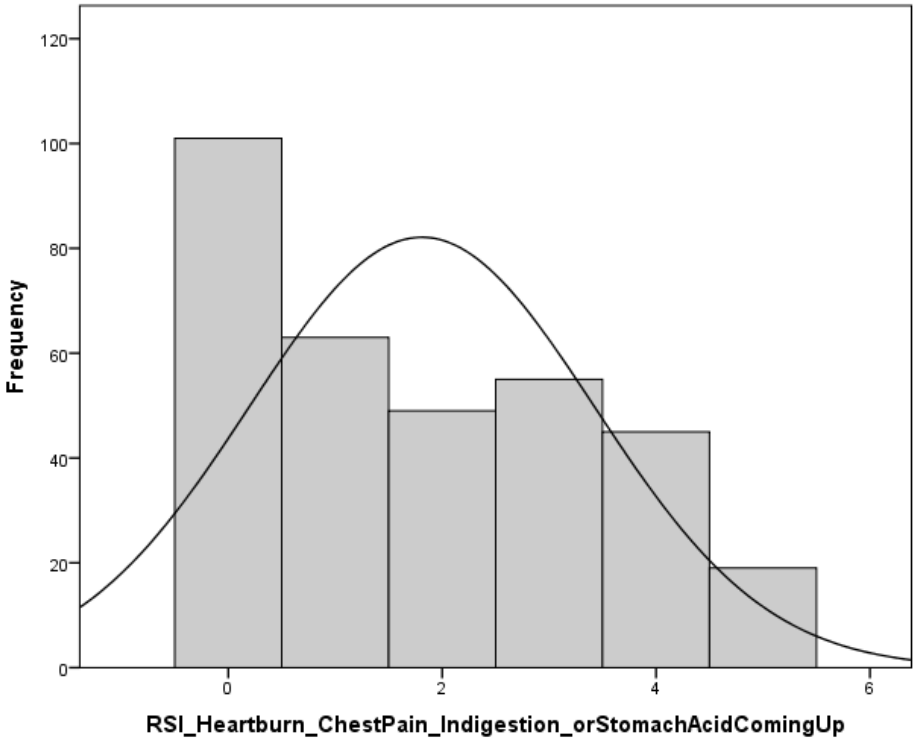


Figure 6-9 Histogram and overlying normal curve showing the distribution of the RSI item 9 - Heartburn



Exploratory Factor Analysis Appendices

EFA Methodology in SPSS

Choose: Analyse / Dimension reduction / Factor analysis

Enter variables (All RSI, CReSS and LPR-HRQL – only the 4 symptom domains)

Descriptives: Choose KMO and Bartlett's test of sphericity. The KMO (Kaiser-Meyer-Olkin measure of sampling adequacy) assesses whether a factor analysis is appropriate for the data and should be >0.6. The higher the value, the more appropriate an EFA is to fit the data. The Bartlett's test of sphericity should by significant (<0.05) to support an EFA

Extraction: The method to choose will depend on the data distribution. Principal Axis Factoring is appropriate for non-normal data and Maximal Likelihood for normal data. These extraction methods are preferable to the SPSS default choice of Principal Components Analysis, as the latter includes both shared and unique variable variance. Principal Axis Factoring and Maximal Likelihood include only shared variable variance in the solution. Chapter 2 showed that the individual items of the RSI, when plotted as a histogram with overlying normal curve (see Chapter 7 Appendices for the graphs), did not appear graphically as normally distributed. Also considering the construction of a Likert scoring scale for items on the questionnaires in which, for example, a rise from 1 to 2 may not necessarily be equivalent to a rise from a score of 3 to 4, it would be appropriate to use the Principal Axis Factoring extraction method for non-normally distributed data.

Display: Scree Plot

Extract: based on Eigenvalues > 1

Extraction of factors can be performed by retaining those with an eigenvalue > 1. However, this may not be accurate. A scree test plots the factors' eigenvalues graphically. The steep aspect of the curve will contain those factors to be retained. After the "break" or "shoulder" of the curve will lie factors contributing less variance. Those factors after the break can be removed. Where there is ambiguity if a factor is at or near the break in the curve, several factor analyses can be run with a differing number of factors to assess which produces the "cleanest pattern matrix". Rather than extracting based on eigenvalues, the number of factors to be included is specified.

Rotation: Promax.

Rotation of the data can either be via an orthogonal (90 degrees) or oblique. Oblique rotation is more appropriate for correlated factors. For this thesis, correlation between throat symptom factors would be expected. This theory would be supported by the strong correlation demonstrated between the questionnaires in Chapter 3. The rotation options in SPSS are Direct Oblimin or Promax. Promax has been recommended for large datasets (James Gaskin reference).

Scores: Nil required

Options: Suppress small coefficients (should be 0.2 below minimum acceptable loading value – eg 0.3 if 0.5 is the desirable minimum item loading to a factor).

EFA Outputs: Check KMO and sphericity significance.

Communalities: the figures in the extraction column depict how an item correlates with all the other items. Individual items should be > 0.2-0.4. Communalities are considered high if all are > 0.8, but that is considered rare. Low to moderate communalities are acceptable in the 0.4 to 0.7 range. The rotation of the model should converge within 10 iterations to denote an acceptable EFA.

Total variance explained – shows how many factors have been extracted and how much of the total variance these factors explain. Use Extraction Sums of Squared Loadings value rather than the Initial Eigenvalue percentage variance.

Assess scree plot and number of factors above 1. The alternative is to take the number of factors above the shoulder of the curve of the scree plot, which may not be the same as Eigenvalue > 1.

Pattern Matrix: Individual items are presented in the left hand column, the factors extracted in the superior row, with the loading value of each item per factor displayed. The factor correlations can be analysed separately, aiming for the separate factors to not correlate with one another – the items in each factor should correlate with each other.

Look for clean factors with no low loading scores (defined in this instance as <0.4):

If one single item loads on more than one factor, this is a cross loading item. This is acceptable if the loadings differ by more than 0.2. Cross loading items can be removed and the EFA re-run, if the factor has several strong loading items (>0.5).

A factor with less than 3 items is weak. Five or more strongly loading items is ideal.

When there is a clean factor structure produced, the individual factors can be named – either according to the specified questions in the tool, or as new latent variable groups.

EFA of three questionnaires with factors defined by Eigenvalues > 1

The following sequence of steps details the items removed in the series of EFA that reduced the initial 15 factor model to a cleaner 10 factor model (Table 7-1).

- 'RSI hoarseness' and 'CReSS hoarseness' cross-loaded with factor 1 and they loaded with the 'CReSS pain in the throat' item in a separate factor. These two items had loading scores >0.4 in both factors. 'CReSS pain in throat' cross-loaded with 'I am afraid of choking in my sleep' and 'I awaken from sleep gasping for breath' symptoms from the LPR HRQL Swallow domain. The 'CReSS Pain in the Throat' item was removed due to this cross-loading.

- The LPR Voice 1 and 11 items cross loaded across two factors, both with scores > 0.5. These items are very similar in description; 'My voice makes it difficult to work' and 'my voice makes it difficult to do my job'. Item 11 was removed and the EFA conducted again.

- The LPR-HRQL Swallow 29 item ('I hesitate to eat in public because I have trouble swallowing') cross-loaded across two factors with low loading scores, 0.41 and 0.28.

- The 'CReSS rush of saliva' cross-loaded across two factors with low loading scores, 0.21 and 0.21.

- The LPR-HRQL Swallow 28 item ('I feel a lump in my throat which makes it hard to swallow') cross loaded with low scores across four factors, and loaded on a separate factor with a number of cough related symptoms from the emerging second factor.

- The 'CReSS hiccups' item cross-loaded with low scores, 0.27 and 0.23.

- The 'CReSS back pain', 'pressure in chest' and 'headache' items all cross loading with low values.

- The LPR-HRQL Swallow 31 item ('I am bothered by burning feeling in my throat') cross-loaded across two factors with low scores, 0.28 and 0.49.

- The LPR-HRQL Cough 15 item ('I avoid social events where my coughing might bother others') cross loaded across 3 factors, with scores of 0.49, 0.36 and 0.38.

- The 'CReSS feeling full early' item cross loaded across 3 factors with scores of 0.24, 0.22 and 0.22.

- The 'CReSS decreased appetite' item cross-loaded with low scores, 0.29 and 0.21.

- The 'CReSS vomiting' item cross-loaded and low scores, 0.34 and 0.24. The 'CReSS nausea' item cross-loaded with low scores, 0.27 and 0.25. Both items were removed before the EFA was repeated.

- At this stage there were 10 items forming the second factor, which appeared to relate to cough. Five of these items loaded on a second common factor with the LPR-HRQL Cough 14 item cross loading over 3 factors, with scores of 0.59, 0.34 and 0.27. This item was therefore removed. This removal split the second cough factor into 2 separate factors.

- The LPR-HRQL Throat 24 item ('The need to clear my throat makes it hard to talk') cross-loaded over two factors, which is understandable from the phraseology covering voice and throat clearing). The loading scores were 0.42 and 0.37. This removal brought the cough related items back into one factor.

- The 'CReSS bad breath' item was a low score on one factor, 0.28. The 'CReSS gurgling stomach' cross loaded across two factors, with low scores 0.46 and 0.24.

Table 7-1 Ten factor EFA model – Pattern Matrix

Questionnaire Item	**Factor Number**									
	1	2	3	4	5	6	7	8	9	10
lprVOICE8	0.90									
lprVOICE3	0.87									
lprVOICE9	0.84									
lprVOICE_4	0.81									
lprVOICE10	0.78									
lprVOICE 1	0.75									
CReSS_Hoarseness	0.74				0.21				-0.27	
RSI_Hoarseness_	0.73								-0.26	
lprVOICE6	0.70								0.25	
lprVOICE12	0.64									
lprVOICE 5	0.61									
lprVOICE 7	0.60									
CReSS_Cough Upright		0.81								
RSI_Cough-after Eating / Lying		0.78								
lprCOUGH 17		0.78								
RSI_Troublesome Cough		0.78				-0.24				
lprCOUGH 19		0.76								
CReSS_CoughLying		0.74								
lprCOUGH 16		0.72								
CReSS_CoughafterEating		0.67								
lprCOUGH 18		0.65								
CReSS_Heartburn			0.92							
RSI_Heartburn_			0.90							

CReSS_StomachAcidUp			0.86							
CReSS_Indigestion			0.79							
CReSS_AcidSourMouth			0.59							
CReSS_Regurgitation			0.44							
CReSS_DifficultyBreathing				0.84						
lprSWALLOW 32				0.63						
lprSWALLOW 30				0.59						
RSI_BreathingDifficulties_orChokingEpisodes				0.56						
CReSS_Wheezing				0.54						
CReSS_Choking				0.54		0.23				
CReSS_ThroatClearing					0.79					
RSI_ThroatClear					0.73					
lprTHROAT 25					0.68				0.28	
lprTHROAT 21					0.60				0.36	
CReSS_DifficultySwalFood						0.91				
RSI_DifficultySwallowingFood_Liquids						0.81			-0.20	
CReSS_DifficultySwalLiquids						0.71				
CReSS_ExcessMucus							0.91			
CReSS_MucusDripping							0.76			
RSI_ExcessThroatMucus_orPND							0.75			
RSI_SomethingCaught_orLumpInThroat								0.90		
CReSS_LumpinThroat								0.67		
CReSS_FeelingThingsStuckThroat								0.67		
lprTHROAT 23									0.68	
lprTHROAT 22									0.58	
lprTHROAT 26									0.52	
CReSS_Belching										0.73
CReSS_Flatulence										0.71

CReSS_Bloating										0.49

Table 7-2 EFA communalities - Combined Questionnaires. seven factors specified

Questionnaire Item	Extraction
RSI Hoarseness or problem with voice	0.58
RSI Throat Clear	0.52
RSI Excess Throat Mucus or PND	0.53
RSI Difficulty Swallowing Food Liquids or Tablets	0.53
RSI Cough after Eating or Lying	0.63
RSI Breathing Difficulties or Choking Episodes	0.41
RSI Troublesome Cough	0.60
RSI Something Caught or Lump In Throat	0.49
RSI Heartburn Chest Pain Indigestion or Stomach Acid Coming Up	0.66
CReSS Heartburn	0.67
CReSS Regurgitation	0.46
CReSS Acid Sour Mouth	0.51
CReSS Lump in Throat	0.50
CReSS Difficulty Swallowing Food	0.64
CReSS Difficulty Swallowing Liquids	0.42
CReSS Bloating	0.31
CReSS Belching	0.38
CReSS Flatulence	0.21
CReSS Choking	0.49
CReSS Cough Upright	0.68
CReSS Cough after Eating	0.57
CReSS Cough Lying	0.62
CReSS Wheezing	0.43
CReSS Difficulty Breathing	0.44
CReSS Hoarseness	0.61
CReSS ThroatClearing	0.62
CReSS ExcessMucus	0.63
CReSS MucusDripping	0.61
CReSS FeelingThingsStuckThroat	0.57
CReSS Indigestion	0.69
CReSS StomachAcidUp	0.72
lprVOICE 1 My Voice Problems Make It Difficult For Me To Work	0.52
lprVOICE 3 Being Hoarse Makes It Hard For Me To Communicate My True Self	0.70
lprVOICE 4 My Voice Makes Others Feel Uncomfortable To Listen To Me	0.67
lprVOICE 5 I Can’t Sing As Much As I Would Like To Because Of My Voice	0.39
lprVOICE 6 I Find It Hard To Meet New People Because Of What They Will Think	0.63
lprVOICE 7 The Sound Of My Voice Makes People Think I’m Angry Or Upset	0.51
lprVOICE 8 Straining To Talk Is Tiring	0.74
lprVOICE 9 I Feel Embarrassed About The Sound Of My Voice	0.74
lprVOICE 10 I Avoid Talking Because Of The Effort	0.67

lprVOICE 12 I Am Afraid I Might Lose My Voice Forever	0.53
lprCOUGH 16 I Have To Leave The Room Because Of My Coughing	0.56
lprCOUGH 17 People Think I Am Sick Because Of My Coughing	0.65
lprCOUGH 18 My Co-workers Can Hear Me Coming Because Of My Coughing	0.52
lprCOUGH 19 I Worry About Having A Coughing Spell At A Bad Time	0.64
lprTHROAT 21 People Notice How Much I Have To Clear My Throat	0.52
lprTHROAT 22 Clearing My Throat Has A Negative Effect On Sex	0.48
lprTHROAT 23 Clearing My Throat Has A Negative Effect On Friendships	0.59
lprTHROAT 25 I Feel Frustrated About Having To Clear My Throat So Often	0.56
lprTHROAT 26 I Avoid Social Events Because Of The Need To Clear My Throat	0.47
lprSWALLOW 30 I Am Afraid Of Choking In My Sleep	0.41
lprSWALLOW 32 I Awaken From Sleep Gasping For Breath	0.43

Table 7-3 CReSS EFA communalities scores

Questionnaire Item	Extraction
CReSS_Heartburn	0.78
CReSS_PressureChest	0.53
CReSS_Regurgitation	0.51
CReSS_Acid SourMouth	0.54
CReSS_Gurgling Stomach	0.48
CReSS_Lump in Throat	0.64
CReSS_Difficulty Swallowing Food	0.75
CReSS_Difficulty Swallowing Liquids	0.60
CReSS_Nausea	0.46
CReSS_Pain in Throat	0.26
CReSS_Vomiting	0.38
CReSS_Bloating	0.43
CReSS_Belching	0.63
CReSS_Flatulence	0.56
CReSS_Hiccups	0.29
CReSS_Decreased Appetite	0.49
CReSS_Rush of Saliva	0.41
CReSS_Feeling Full Early	0.54
CReSS_Bad Breath	0.28
CReSS_Back Pain	0.25
CReSS_Headache	0.25
CReSS_Choking	0.40
CReSS_Cough Upright	0.78
CReSS_Cough after Eating	0.63
CReSS_Cough Lying	0.61
CReSS_Wheezing	0.59
CReSS_Difficulty Breathing	0.54
CReSS_Hoarseness	0.22
CReSS_Throat Clearing	0.42

CReSS_Excess Mucus	0.75
CReSS_Mucus Dripping	0.66
CReSS_Feeling Things Stuck Throat	0.56
CReSS_Indigestion	0.72
CReSS_Stomach Acid coming Up	0.74

Figure 6-10 CReSS EFA Scree Plot

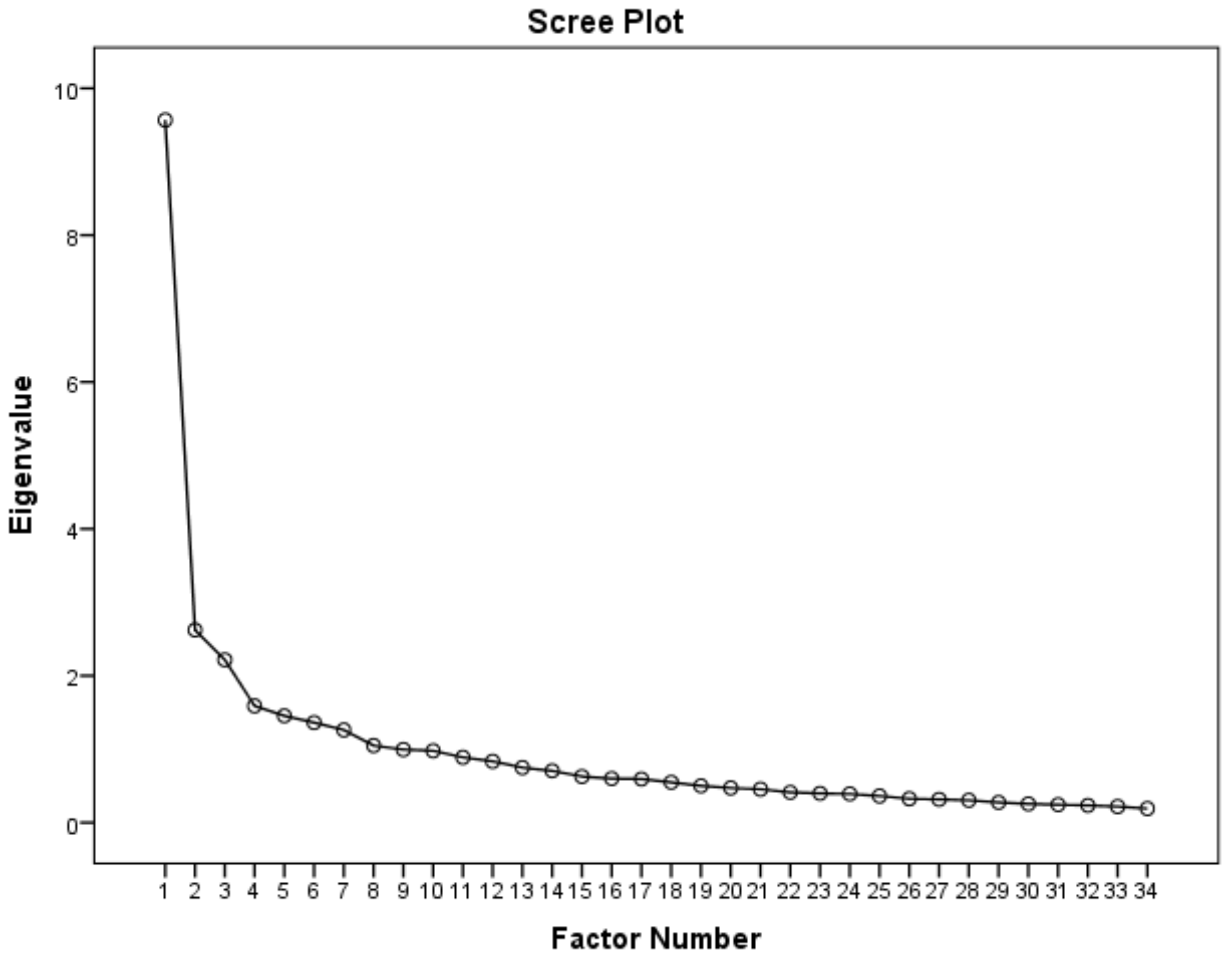


Table 7-4 CReSS EFA Pattern Matrix

Questionnaire Item	Factor							
	1	2	3	4	5	6	7	8
CReSS_Heartburn	0.95							
CReSS_Stomach acid coming up	0.86							
CReSS_Indigestion	0.81							
CReSS_Acid or sour taste in Mouth	0.46	0.33						
CReSS_Vomiting		0.68				-0.23		
CReSS_Feeling Full Early		0.65	0.22					
CReSS_Decreased Appetite		0.62						
CReSS_Nausea		0.50						
CReSS_Hiccups		0.43						
CReSS_Headache		0.41						
CReSS_Rush of saliva in mouth		0.40						
CReSS_Regurgitation	0.35	0.40						
CReSS_Bad Breath		0.39						

CReSS_Difficulty breathing			0.70					
CReSS_Wheezing			0.64	0.22				
CReSS_Pressure in chest			0.62				0.23	
CReSS_Hoarseness			0.34					
CReSS_Back pain		0.24	0.31					
CReSS_Coughing when upright				0.88				
CReSS_Coughing after eating				0.74				
CReSS_Coughing lying down				0.72				
CReSS_Excess mucus					0.92			
CReSS_Mucus dripping in throat					0.77			
CReSS_Throat clearing					0.58	0.20		
CReSS_Flatulence						0.82		
CReSS_Belching						0.71		
CReSS_Gurgling stomach						0.51		0.21
CReSS_Bloating						0.47		
CReSS_Lump in throat							0.77	

CReSS_Feeling things stuck in throat							0.64	
CReSS_Pain in throat							0.30	
CReSS_Difficulty swallowing food								0.83
CReSS_Difficulty swallowing liquids								0.73
CReSS_Choking		0.21		0.24				0.29

Table 7-5 CReSS three Factor EFA Pattern Matrix

Questionnaire Item	Factor		
	1	2	3
CReSS_Indigestion	0.92		
CReSS_Stomach acid coming up	0.88		
CReSS_Heartburn	0.82		
CReSS_Acid or sour taste in mouth	0.67		
CReSS_Regurgitation	0.61		
CReSS_Belching	0.60		
CReSS_Bloating	0.56		
CReSS_Gurgling stomach	0.51		
CReSS_Flatulence	0.41		
CReSS_Nausea	0.38		0.30
CReSS_Feeling full early	0.37		0.31
CReSS_Rush of saliva	0.34	0.23	
CReSS_Hiccups	0.27		
CReSS_Bad breath	0.25		
CReSS_Coughing when upright		0.88	
CReSS_Coughing when lying		0.75	
CReSS_Cough after eating		0.72	
CReSS_Wheezing		0.64	
CReSS_Mucus dripping in throat	0.28	0.52	
CReSS_Excess mucus	0.26	0.52	
CReSS_Throat clearing		0.45	
CReSS_Difficulty breathing		0.45	0.26
CReSS_Hoarseness		0.33	
CReSS_Back pain		0.24	
CReSS_Difficulty swallowing food			0.84
CReSS_Difficulty swallowing liquids			0.73
CReSS_Lump in throat			0.60
CReSS_Feeling things stuck in throat			0.54
CReSS_Decreased appetite			0.51
CReSS_Pain in throat			0.50
CReSS_Choking		0.32	0.47
CReSS_Headache			0.36
CReSS_Pressure chest	0.22		0.31

CReSS_Vomiting	0.21		0.23

Cluster Analysis Appendices

Figure 6-11 Two-step cluster analysis predictor importance for seven factor model with unspecified cluster number

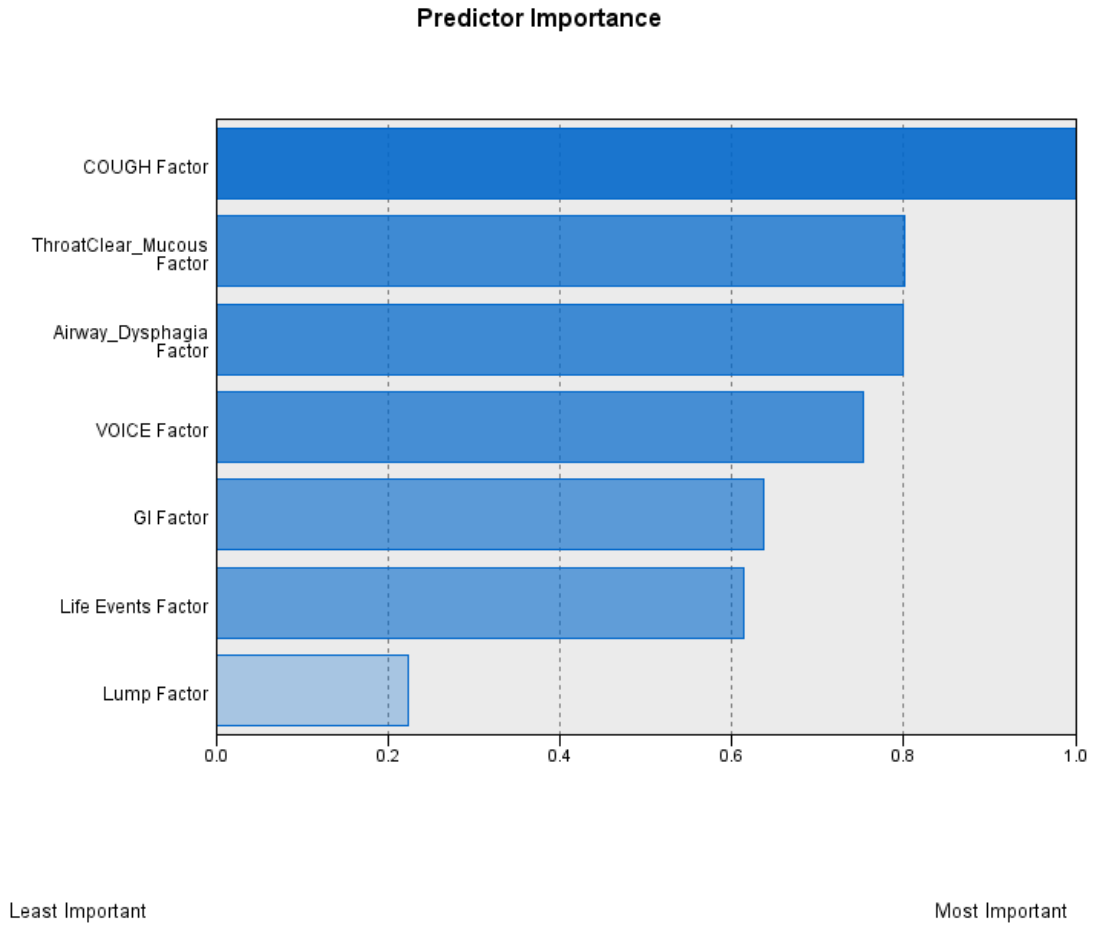
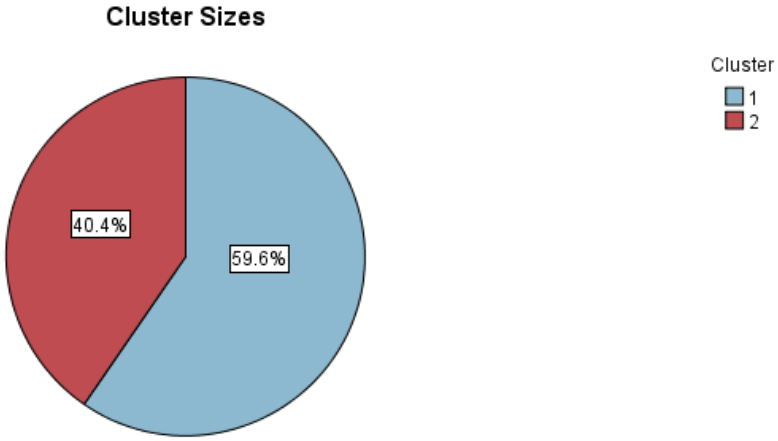


Figure 6-12 Two-step cluster analysis - Cluster sizes for seven factor model with unspecified cluster numbers



Size of Smallest Cluster	134 (40.4%)
Size of Largest Cluster	198 (59.6%)
Ratio of Sizes: Largest Cluster to Smallest Cluster	1.48

Figure 6-13 Two step cluster analysis with number of clusters undefined

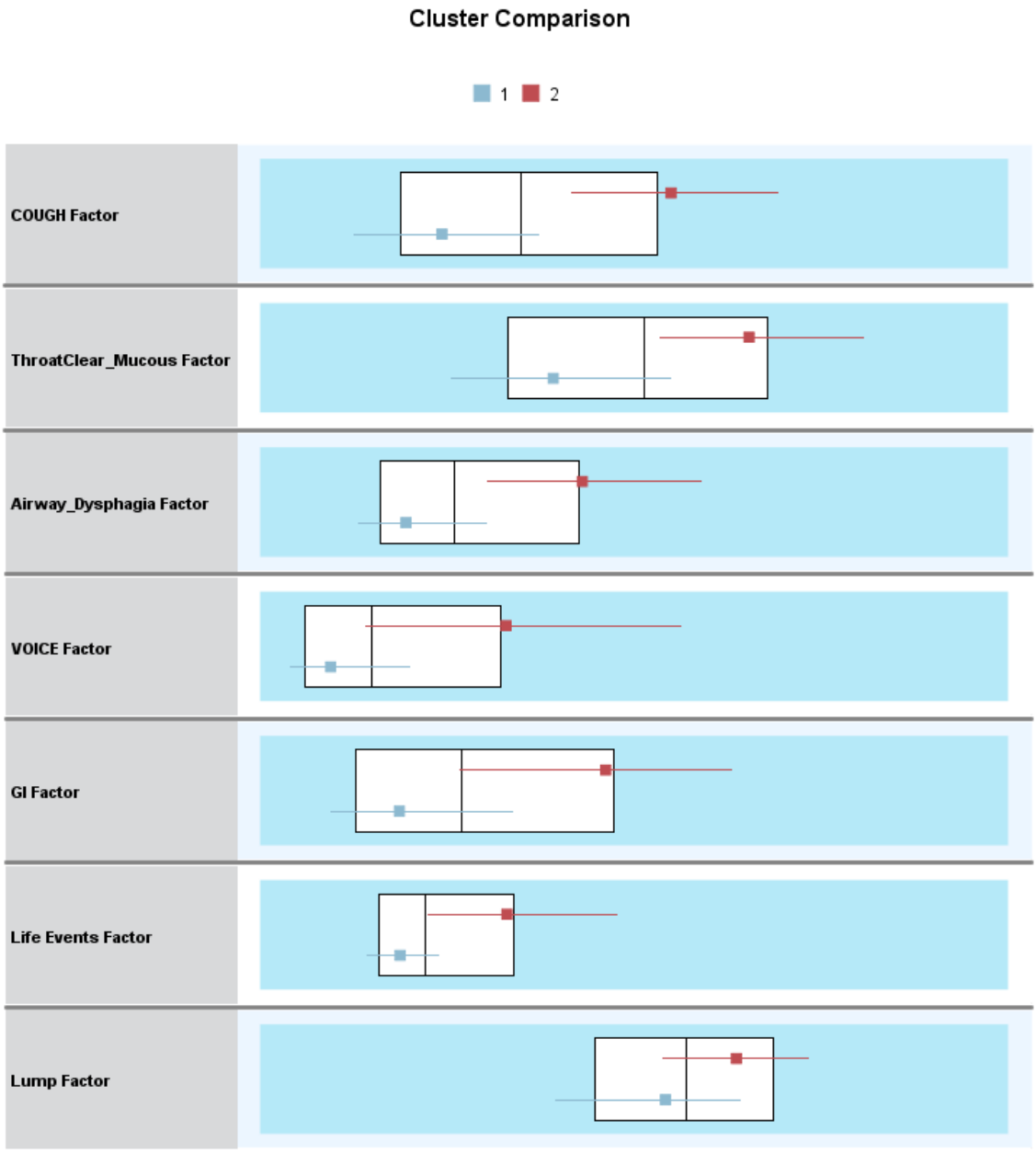


Figure 6-14 Bar chart to show two cluster distribution for seeven factors with K-means analysis

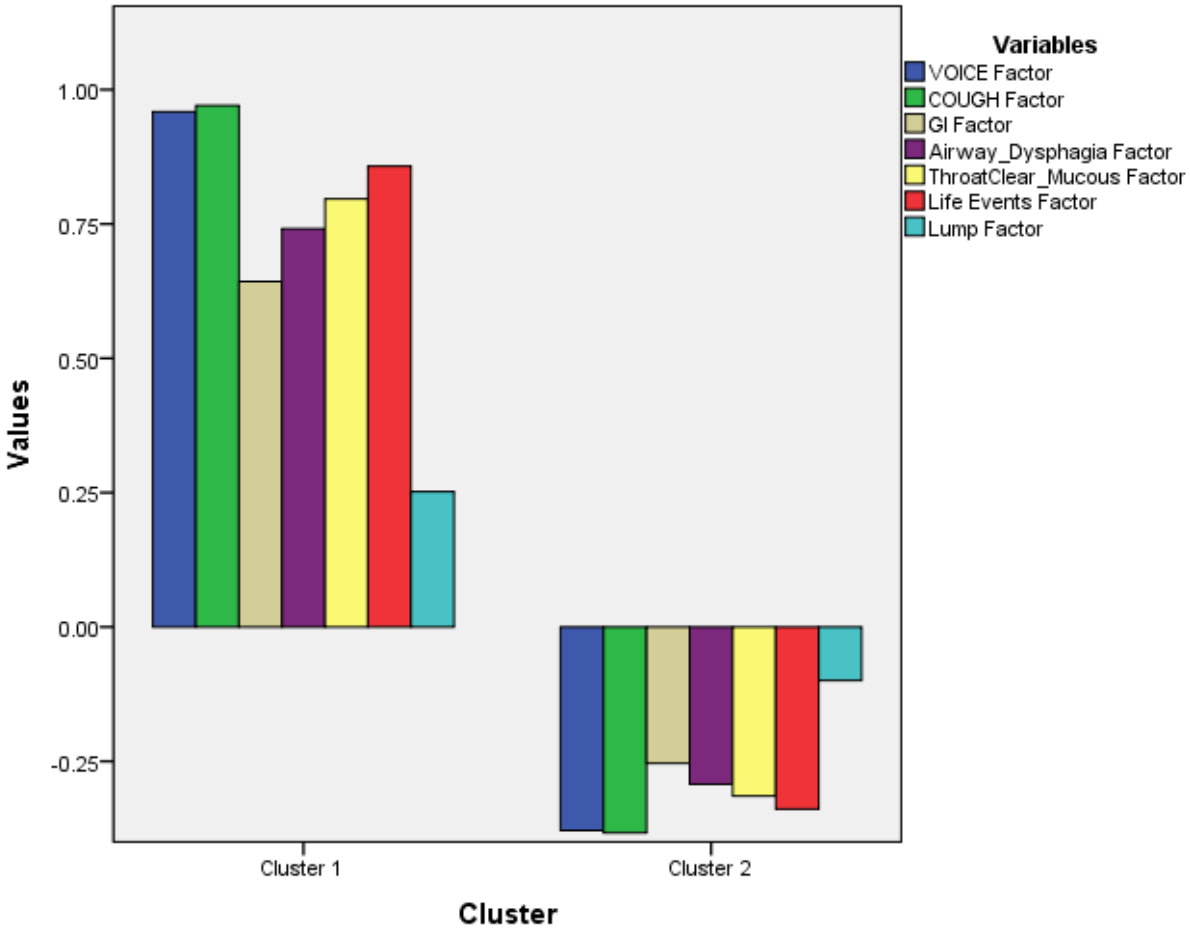
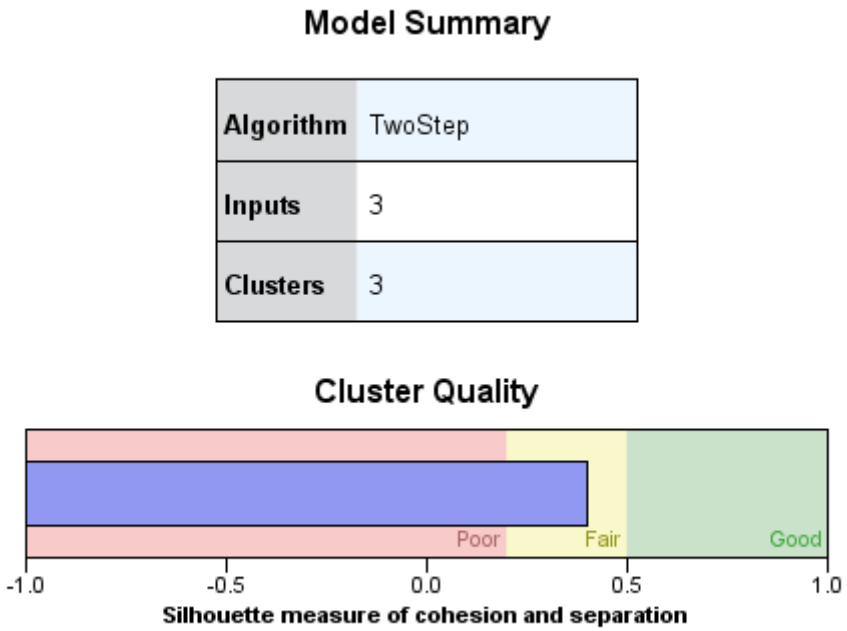


Table 7-6 ANOVA for seven factor two cluster analysis

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
VOICE Factor	120.54	1	0.59	330	203.87	.000
COUGH Factor	123.49	1	0.56	330	219.26	.000
GI Factor	54.20	1	0.76	330	71.25	.000
Airway_Dysphagia Factor	72.04	1	0.66	330	109.41	.000
ThroatClear_Mucus Factor	83.25	1	0.65	330	128.66	.000
Life Events Factor	96.40	1	0.55	330	175.90	.000
Lump Factor	8.32	1	0.77	330	10.83	.001

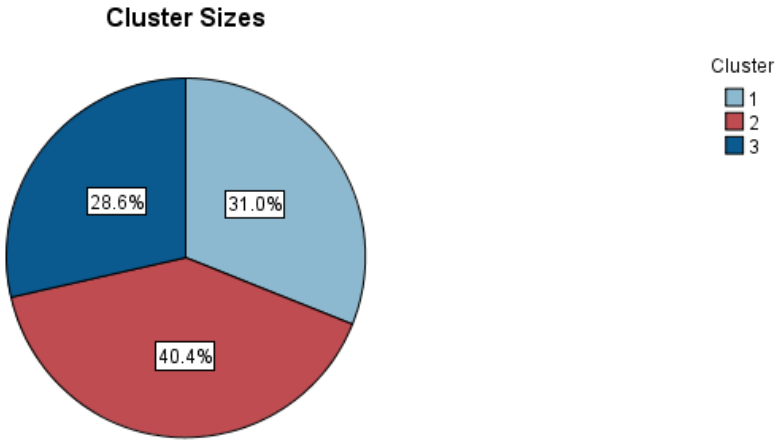
df, degrees of freedom; F, F-score, Sig, statistical significance

Figure 6-15 Model summary: Two-step cluster analysis with three factors, three .



.

Figure 6-16 Cluster sizes - Two-step cluster analysis with three factors, three clusters



Size of Smallest Cluster	95 (28.6%)
Size of Largest Cluster	134 (40.4%)
Ratio of Sizes: Largest Cluster to Smallest Cluster	1.41

Figure 6-17 Predictor importance: Two-step cluster analysis with three factors, three clusters specified.

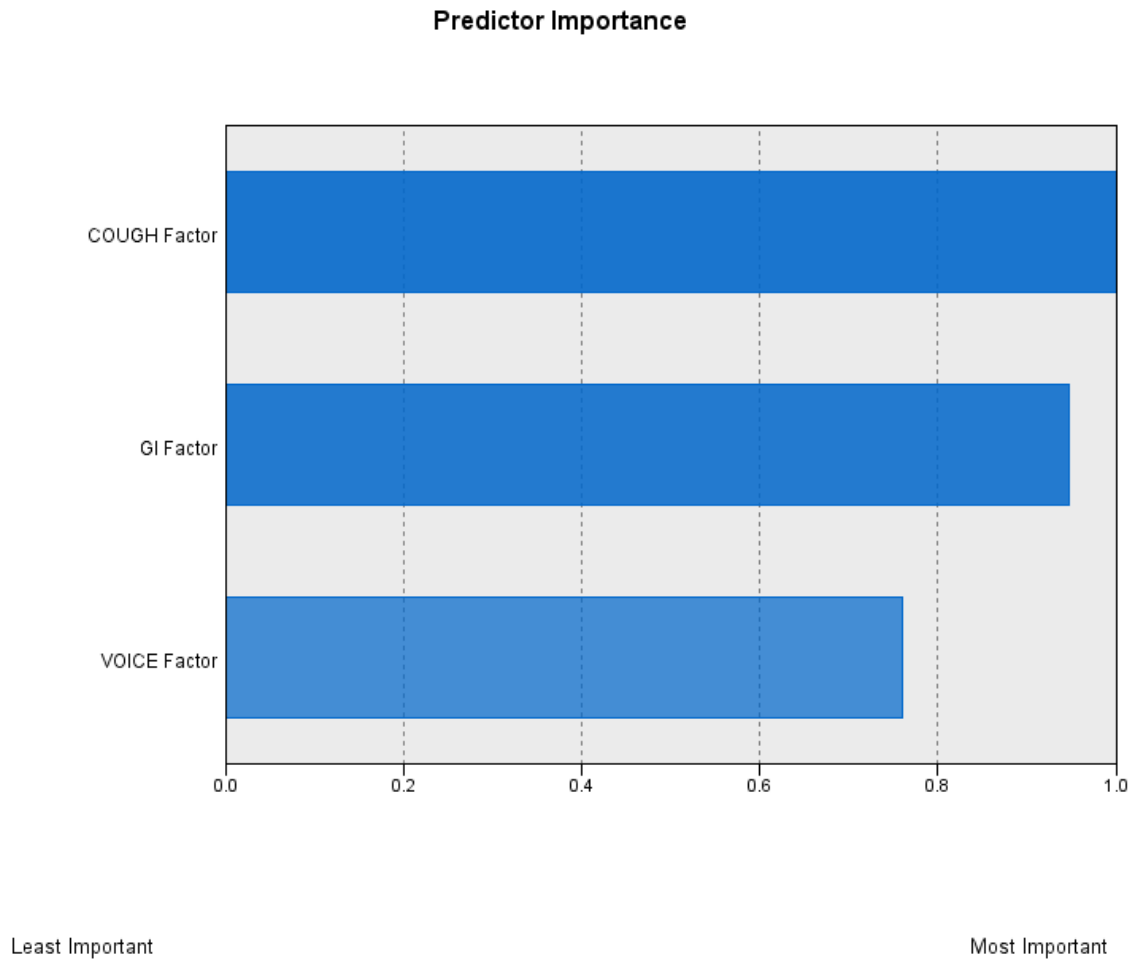


Figure 6-18 Two-step cluster analysis with three factors, four clusters specified.

Model Summary

Algorithm	TwoStep
Inputs	3
Clusters	4

Cluster Quality

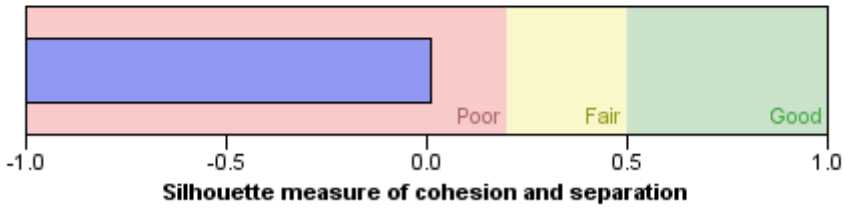
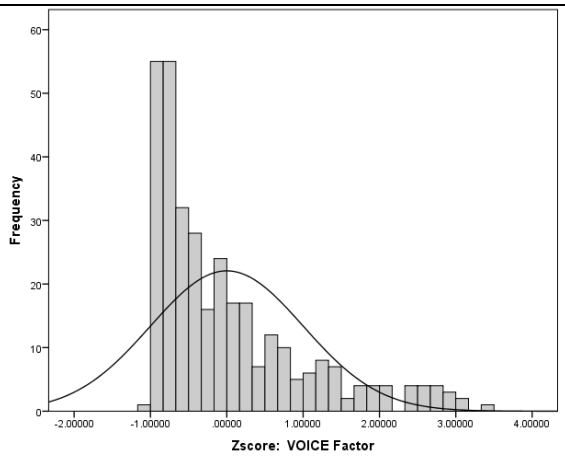
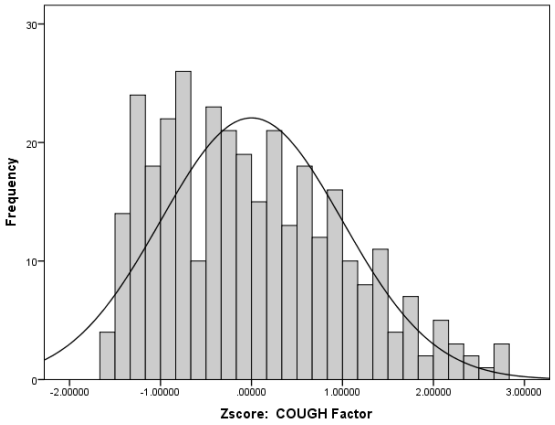
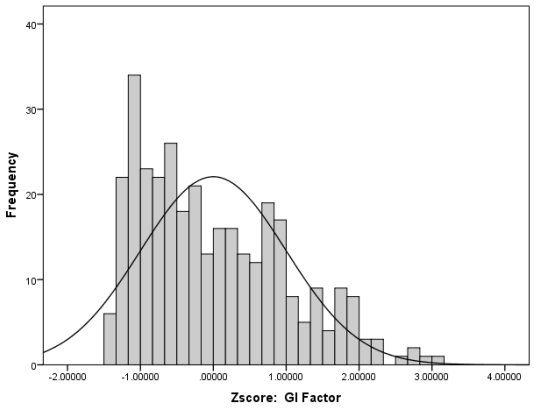
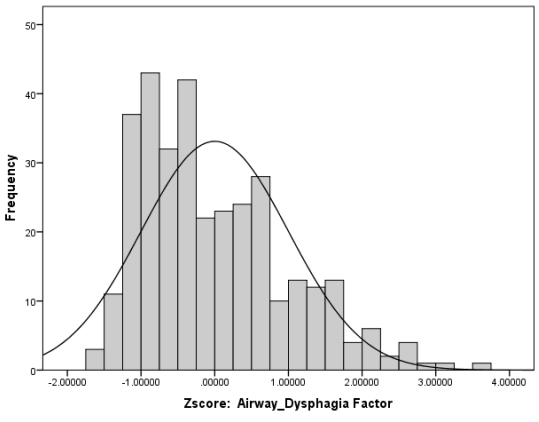
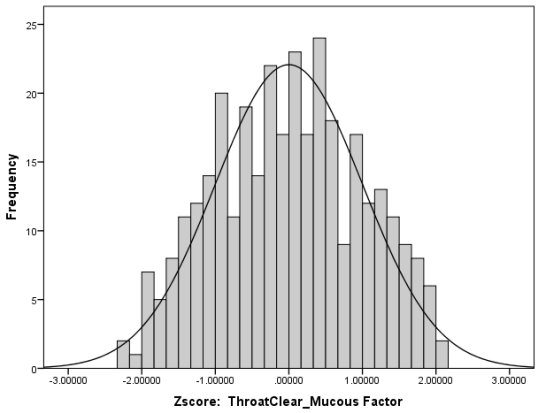


Table 7-7 Histograms with overlying normal curves, Skewness and Kurtosis scores for standardised factor scores

Voice Skewness 1.39 Kurtosis 1.25	**Cough** Skewness 0.53 Kurtosis -0.45
	
GI Skewness 0.68 Kurtosis -0.33	**Airway and Dysphagia** Skewness 0.84 Kurtosis 0.23
	

Throat Clear and Mucus Skewness -0.01 Kurtosis -0.78	**Life Events** Skewness 1.55 Kurtosis 2.88
	
Lump Factor Skewness -0.15 Kurtosis -0.22	

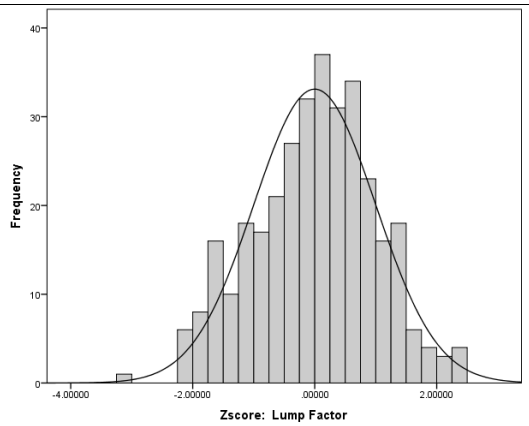
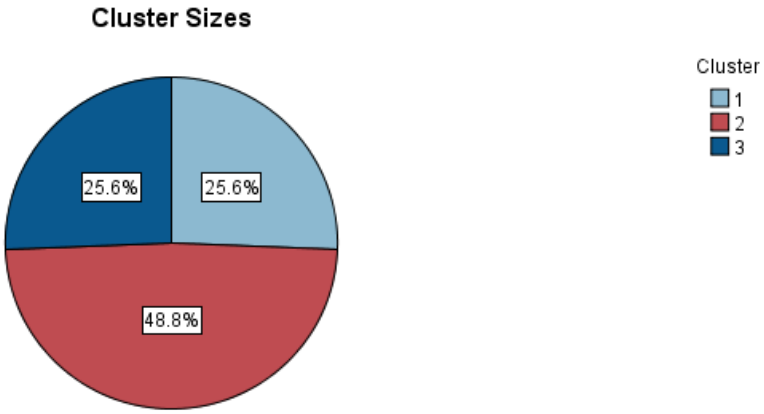


Table 7-8 ANOVA table for K means cluster analysis, seven factor, three clusters

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
VOICE Factor	49.31	2	0.66	329	74.75	.000
COUGH Factor	80.41	2	0.45	329	178.09	.000
GI Factor	47.11	2	0.64	329	73.46	.000
Airway_Dysphagia Factor	45.80	2	0.60	329	76.21	.000
ThroatClear_Mucus Factor	48.89	2	0.61	329	80.82	.000
Life Events Factor	61.88	2	0.47	329	132.63	.000
Lump Factor	6.83	2	0.75	329	9.06	.000

df, degrees of freedom; F, F-score, Sig, statistical significance

Figure 6-19 Two-step cluster analysis RSI, three clusters - Cluster sizes



Size of Smallest Cluster	85 (25.6%)
Size of Largest Cluster	162 (48.8%)
Ratio of Sizes: Largest Cluster to Smallest Cluster	1.91

Figure 6-20 Two-step cluster analysis RSI, three clusters - Predictor Importance

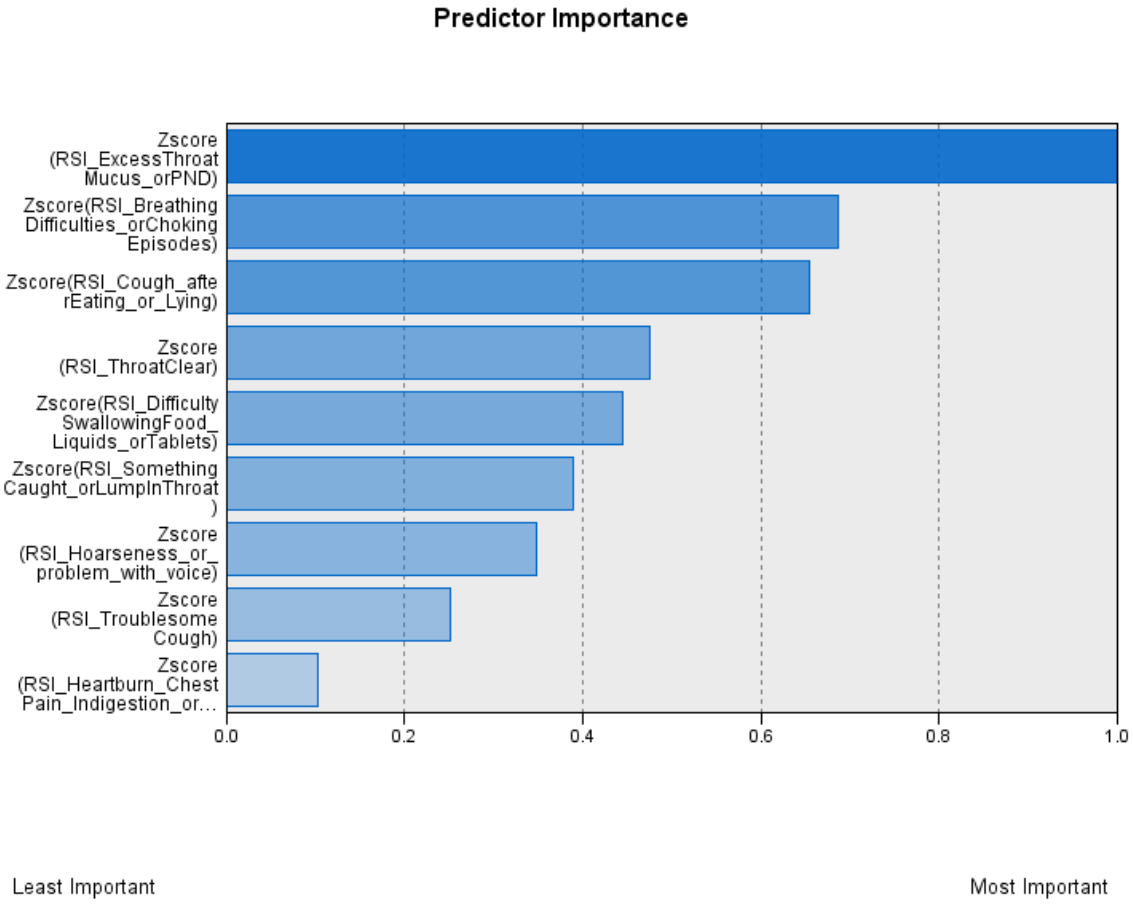


Table 7-9 Cluster 1: hierarchical cluster analysis for the RSI, three clusters specified

	n=	Minimum	Maximum	Mean	Std. Dev.
RSI Hoarseness or problem with voice	138	-1.48	1.59	-.23	.95
RSI Throat Clear	138	-2.68	1.20	-.56	1.06
RSI Excess Throat Mucus or PND	138	-1.84	1.37	-.57	.98
RSI Difficulty Swallowing Food Liquids or Tablets	138	-1.04	2.06	.13	.96
RSI Cough after Eating or Lying	138	-1.32	1.74	-.53	.88
RSI Breathing Difficulties or Choking Episodes	138	-.99	2.09	-.39	.85
RSI Troublesome Cough	138	-1.46	1.46	-.56	.90
RSI Something Caught or Lump In Throat	138	-2.44	1.03	.06	.82
RSI Heartburn Chest Pain. Indigestion or Stomach Acid Coming Up	138	-1.12	1.98	-.34	.88

Std. Dev, standard deviation

Table 7-10 Cluster 2: hierarchical cluster analysis for the RSI, three clusters specified

	n=	Minimum	Maximum	Mean	Std. Dev.
RSI Hoarseness or problem with voice	98	-1.48	1.59	-.35	.98
RSI Throat Clear	98	-2.68	1.20	.20	.80
RSI Excess Throat Mucus or PND	98	-1.84	1.37	.45	.68
RSI Difficulty Swallowing Food Liquids or Tablets	98	-1.04	2.06	-.70	.62
RSI Cough after Eating or Lying	98	-1.32	1.74	-.03	.95
RSI Breathing Difficulties or Choking Episodes	98	-.99	2.09	-.26	.86
RSI Troublesome Cough	98	-1.46	1.46	.23	.90
RSI Something Caught or Lump In Throat	98	-2.44	1.03	-.56	1.26
RSI Heartburn Chest Pain. Indigestion or Stomach Acid Coming Up	98	-1.12	1.98	.31	1.10

Std. Dev, standard deviation

Table 7-11 Cluster 3: hierarchical cluster analysis for the RSI, three clusters specified

	n=	Minimum	Maximum	Mean	Std. Dev.
RSI Hoarseness or problem with voice	96	-1.48	1.59	.69	.71
RSI Throat Clear	96	-1.13	1.20	.60	.59
RSI Excess Throat Mucus or PND	96	-1.84	1.37	.37	.90
RSI Difficulty Swallowing Food Liquids or Tablets	96	-1.04	2.06	.53	.97
RSI Cough after Eating or Lying	96	-.71	1.74	.79	.65
RSI Breathing Difficulties or Choking Episodes	96	-.99	2.09	.83	.84
RSI Troublesome Cough	96	-1.46	1.46	.58	.80
RSI Something Caught or Lump In Throat	96	-1.05	1.03	.48	.59
RSI Heartburn Chest Pain. Indigestion or Stomach Acid Coming Up	96	-1.12	1.98	.17	.92

Std. Dev, standard deviation

Bibliography

The methodology used in the thesis followed instructions from a number of statistical references. Where appropriate, individual statements are referenced. However, given the broad theme of methods described, there is considerable overlap between sources used.

Altman DG. Practical Statistics for Medical Research. 1992. Published by Chapman and Hall, London.

Costello AB., Osborne JW (2005). Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis. Practical Assessment Research and Evaluation; 10(7). http://pareonline.net/getvn.asp?v=10&n=7

Encyclopaedia of Biostatistics. Second Edition. Eds. Armitage & Colton. 2005. Published by Wiley, West Sussex, England.

Factor Analysis and Related Techniques. Ed. M.S.Lewis-Beck. 1994. Published by Sage Publications Ltd.[75]

Gaskin, J (2016), Gaskination's StatWiki. http://statwiki.kolobkreations.com . A series of accessible wed-based tutorial videos were useful in clarifying the approaches in SPSS to conduct the EFA and cluster analysis[68, 72-74].

References

1. Digital N. Hospital Outpatient Activity 2010 - 11. In Digital N (ed). https://digital.nhs.uk/catalogue/PUB03032: 2011.
2. Deary IJ, Wilson JA, Harris MB, MacDougall G. Globus pharyngis: development of a symptom assessment scale. Journal of Psychosomatic Research 1995; 39: 203-213.
3. Lowden M, McGlashan JA, Steel A et al. Prevalence of symptoms suggestive of extra-oesophageal reflux in a general practice population in the UK. Logopedics, Phoniatrics, Vocology 2009; 34: 32-35.
4. Powell J, O'Hara J, Wilson JA. Are persistent throat symptoms atypical features of gastric reflux and should they be treated with proton pump inhibitors? BMJ 2014; 349: g5813.
5. O'Hara J, Miller T, Carding P et al. Relationship between fatigue, perfectionism, and functional dysphonia. Otolaryngology - Head & Neck Surgery 2011; 144: 921-926.
6. MacKenzie K, Millar A, Wilson JA et al. Is voice therapy an effective treatment for dysphonia? A randomised controlled trial. BMJ 2001; 323: 658-661.
7. Excellence NIfHaC. Referral Guidelines for suspected cancer. In Excellence NIfHaC (ed). London: 2015.
8. Thompson WG, Heaton KW. Heartburn and globus in apparently healthy people. Canadian Medical Association Journal 1982; 126: 46-48.
9. Cashman EC, Donnelly MJ. The natural history of globus pharyngeus. International journal of otolaryngology 2010; 2010: 159630.
10. Purcell J. A Treatise of Vapours or Hysterick Fits. Edward Place, London. UK,1707.
11. Malcomson KG. Globus hystericus vel pharyngis (a recommaissance of proximal vagal modalities). Journal of Laryngology & Otology 1968; 82: 219-230.
12. Doody J, Fenton JE. Troublesome Throat Awareness (tTA) as a contemporary alternative to 'globus pharyngeus'. Surgeon Journal of the Royal Colleges of Surgeons of Edinburgh & Ireland 2017; 15: 183-185.
13. O'Hara J, Jones NS. "Post-nasal drip syndrome": most patients with purulent nasal secretions do not complain of chronic cough. Rhinology 2006; 44: 270-273.
14. O'Hara J, Jones NS. The aetiology of chronic cough: a review of current theories for the otorhinolaryngologist. Journal of Laryngology & Otology 2005; 119: 507-514.
15. Harris MB, Deary IJ, Wilson JA. Life events and difficulties in relation to the onset of globus pharyngis. Journal of Psychosomatic Research 1996; 40: 603-615.
16. Deary IJ, Scott S, Wilson IM et al. Personality and psychological distress in dysphonia. British Journal of Health Psychology 1997; 2: 333-341.
17. Gale CR, Wilson JA, Deary IJ. Globus sensation and psychopathology in men: The Vietnam experience study. Psychosomatic Medicine 2009; 71: 1026-1031.
18. Dent J, El-Serag HB, Wallander MA, Johansson S. Epidemiology of gastro-oesophageal reflux disease: a systematic review. Gut 2005; 54: 710-717.
19. Bulmer DM, Ali MS, Brownlee IA et al. Laryngeal mucosa: its susceptibility to damage by acid and pepsin. Laryngoscope 2010; 120: 777-782.
20. Wood JM, Hussey DJ, Woods CM et al. Biomarkers and laryngopharyngeal reflux. Journal of Laryngology & Otology 2011; 125: 1218-1224.
21. Koufman JA. The otolaryngologic manifestations of gastroesophageal reflux disease (GERD): a clinical investigation of 225 patients using ambulatory 24-hour pH monitoring and an experimental investigation of the role of acid and pepsin in the development of laryngeal injury. Laryngoscope 1991; 101: 1-78.
22. Belafsky PC, Postma GN, Koufman JA. Validity and reliability of the reflux symptom index (RSI). Journal of Voice 2002; 16: 274-277.
23. Belafsky PC, Postma GN, Koufman JA. The validity and reliability of the reflux finding score (RFS). Laryngoscope 2001; 111: 1313-1317.
24. Karkos PD, Benton J, Leong SC et al. Trends in laryngopharyngeal reflux: a British ENT survey. European Archives of Oto-Rhino-Laryngology 2007; 264: 513-517.

25. Karkos PD, Wilson JA. Empiric treatment of laryngopharyngeal reflux with proton pump inhibitors: a systematic review. Laryngoscope 2006; 116: 144-148.
26. Mesallam TA, Stemple JC, Sobeih TM, Elluru RG. Reflux symptom index versus reflux finding score. Annals of Otology, Rhinology & Laryngology 2007; 116: 436-440.
27. Lee YS, Choi SH, Son YI et al. Prospective, observational study using rabeprazole in 455 patients with laryngopharyngeal reflux disease. European Archives of Oto-Rhino-Laryngology 2011; 268: 863-869.
28. Habermann W, Schmid C, Neumann K et al. Reflux Symptom Index and Reflux Finding Score in Otolaryngologic Practice. Journal of Voice 2012; 26: e123-e127.
29. Merati AL, Lim HJ, Ulualp SO, Toohill RJ. Meta-analysis of upper probe measurements in normal subjects and patients with laryngopharyngeal reflux. Annals of Otology, Rhinology & Laryngology 2005; 114: 177-182.
30. Schneider GT, Vaezi MF, Francis DO. Reflux and Voice Disorders: Have We Established Causality? Current Otorhinolaryngology Reports 2016; 4: 157-167.
31. Koufman JA. Low-acid diet for recalcitrant laryngopharyngeal reflux: therapeutic benefits and their implications. Annals of Otology, Rhinology & Laryngology 2011; 120: 281-287.
32. McGlashan JA, Johnstone LM, Sykes J et al. The value of a liquid alginate suspension (Gaviscon Advance) in the management of laryngopharyngeal reflux. European Archives of Oto-Rhino-Laryngology 2009; 266: 243-251.
33. Chappity P, Kumar R, Deka RC et al. Proton Pump Inhibitors Versus Solitary Lifestyle Modification in Management of Laryngopharyngeal Reflux and Evaluating Who is at Risk: Scenario in a Developing Country. Clinical Medicine Insights Ear Nose & Throat 2014; 7: 1-5.
34. Guo H, Ma H, Wang J. Proton Pump Inhibitor Therapy for the Treatment of Laryngopharyngeal Reflux: A Meta-Analysis of Randomized Controlled Trials. Journal of Clinical Gastroenterology 2016; 50: 295-300.
35. Liu C, Wang H, Liu K. Meta-analysis of the efficacy of proton pump inhibitors for the symptoms of laryngopharyngeal reflux. Brazilian Journal of Medical & Biological Research 2016; 49: 04.
36. Wei C. A meta-analysis for the role of proton pump inhibitor therapy in patients with laryngopharyngeal reflux. European Archives of Oto-Rhino-Laryngology 2016; 273: 3795-3801.
37. Megwalu UC. A systematic review of proton-pump inhibitor therapy for laryngopharyngeal reflux. Ear, Nose, & Throat Journal 2013; 92: 364-371.
38. Sen P, Georgalas C, Bhattacharyya AK. A systematic review of the role of proton pump inhibitors for symptoms of laryngopharyngeal reflux. Clinical Otolaryngology 2006; 31: 20-24; discussion 24.
39. Shaheen NJ CS, Bright SD, Madanick RD, Buckmire R, Couch M, Dellon ES, Galanko JA, Sharpless G, Morgan DR, Spacek MB, Heidt-Davis P, Henke D. Aliment Pharmacol Ther 2011; 33: 225-234.
40. Lechien JR, Saussez S, Schindler A et al. Clinical outcomes of laryngopharyngeal reflux treatment: A systematic review and meta-analysis. Laryngoscope 2018; 30: 30.
41. Ezzat WF FS, Fathey H, El Demerdash A. Virtue of adding prokinetics to proton pump inhibitors in the treatment of laryngopharyngeal reflux disease: prospective study. J Otolaryngol Head Neck Surg. 2011; 40: 350-356.
42. Ah-See K, Wilson J. Extra-oesophageal reflux: state of the knowledge base. Clinical Otolaryngology 2012; 37: 9-16.
43. Watson G, O'Hara J, Carding P et al. TOPPITS: Trial Of Proton Pump Inhibitors in Throat Symptoms. Study protocol for a randomised controlled trial. Trials [Electronic Resource] 2016; 17: 175.
44. Papakonstantinou L, Leslie P, Gray J et al. Laryngopharyngeal reflux: a prospective analysis of a 34 item symptom questionnaire. Clinical Otolaryngology 2009; 34: 455-459.
45. Drinnan M, Powell J, Nikkar-Esfahani A et al. Gastroesophageal and extraesophageal reflux symptoms: similarities and differences. Laryngoscope 125: 424-430.
46. Francis DO, McPheeters ML, Noud M et al. Checklist to operationalize measurement characteristics of patient-reported outcome measures. Systematic Reviews 2016; 5: 129.

47. Stanghellini V, Armstrong D, Monnikes H, Bardhan KD. Systematic review: do we need a new gastro-oesophageal reflux disease questionnaire? Alimentary Pharmacology & Therapeutics 2004; 19: 463-479.
48. Carrau RL, Khidr A, Gold KF et al. Validation of a quality-of-life instrument for laryngopharyngeal reflux. Arch Otolaryngol Head Neck Surg 2005; 131: 315-320.
49. Printza A, Kyrgidis A, Oikonomidou E, Triaridis S. Assessing laryngopharyngeal reflux symptoms with the Reflux Symptom Index: validation and prevalence in the Greek population. Otolaryngology - Head & Neck Surgery 2011; 145: 974-980.
50. Cathcart RA, Steen N, Natesh BG et al. Non-voice-related throat symptoms: comparative analysis of laryngopharyngeal reflux and globus pharyngeus scales. Journal of Laryngology & Otology 2011; 125: 59-64.
51. Adnane C, Adouly T, Khallouk A et al. Using preoperative unsupervised cluster analysis of chronic rhinosinusitis to inform patient decision and endoscopic sinus surgery outcome. European Archives of Oto-Rhino-Laryngology 2017; 274: 879-885.
52. Powell J, Cocks HC. Mucosal changes in laryngopharyngeal reflux--prevalence, sensitivity, specificity and assessment. Laryngoscope 2013; 123: 985-991.
53. Belafsky PC, Postma GN, Koufman JA. Validity and reliability of the reflux symptom index (RSI). J Voice 2002; 16: 274-277.
54. Papakonstantinou L, Leslie P, Gray J et al. Laryngopharyngeal reflux: a prospective analysis of a 34 item symptom questionnaire. Clin Otolaryngol 2009; 34: 455-459.
55. Vaezi MF, Richter JE, Stasney CR et al. Treatment of chronic posterior laryngitis with esomeprazole. Laryngoscope 2006; 116: 254-260.
56. Reichel O, Dressel H, Wiederanders K, Issing WJ. Double-blind, placebo-controlled trial with esomeprazole for symptoms and signs associated with laryngopharyngeal reflux. Otolaryngology - Head & Neck Surgery 2008; 139: 414-420.
57. Carrau RL, Khidr A, Gold KF et al. Validation of a quality-of-life instrument for laryngopharyngeal reflux. Archives of Otolaryngology -- Head & Neck Surgery 2005; 131: 315-320.
58. Andersson O, Ryden A, Ruth M et al. Development and validation of a laryngopharyngeal reflux questionnaire, the Pharyngeal Reflux Symptom Questionnaire. Scandinavian Journal of Gastroenterology 2010; 45: 147-159.
59. Lee JS, Lee YC, Kim SW et al. Changes in the quality of life of patients with laryngopharyngeal reflux after treatment. Journal of Voice 2014; 28: 487-491.
60. Lechien JR, Finck C, Huet K et al. Impact of age on laryngopharyngeal reflux disease presentation: a multi-center prospective study. European Archives of Oto-Rhino-Laryngology 2017; 274: 3687-3696.
61. Jacobson BC, Somers SC, Fuchs CS et al. Body-mass index and symptoms of gastroesophageal reflux in women. New England Journal of Medicine 2006; 354: 2340-2348.
62. Sone M, Katayama N, Kato T et al. Prevalence of laryngopharyngeal reflux symptoms: comparison between health checkup examinees and patients with otitis media. Otolaryngology - Head & Neck Surgery 2012; 146: 562-566.
63. JD E. Straightforward Statistics for the Behavioural Sciences. Pacific Grove, California: Brooks/Cole Publishing,1996.
64. Obesity. In. https://www.nhs.uk/common-health-questions/lifestyle/what-is-the-body-mass-index-bmi/: NHS 2019.
65. Costello AB OJ. Best practices in exploratory factor analysis: four

recommendations for getting the most from your analysis. Practical Assessment Research & Evaluation, Vol 10, No 7 2005; 10: 1-9.
66. J. G. Exploratory Factor Analysis. In Gaskination's StatWiki

http://statwiki.kolobkreations.com: 2016.
67. Gaskin J. SEM Series Part 3: Exploratory Factor Analysis. In Gaskination's Statistics. http://youtube.com/Gaskination: 2013.

68. Gaskin J. SEM Series (2016) 3. Exploratory Factor Analysis (EFA). In Gaskination's Statistics. http://youtube.com/Gaskination: 2016.
69. Gaskin J. Working through a messy EFA (factor analysis) in SPSS. In Gaskination's Statistics. http://youtube.com/Gaskination: 2016.
70. Daniel J. Mundfrom DGS, Tian Lu Ke. Minimum Sample Size Recommendations for Conducting Factor Analyses. International Journal of Testing 2005; 5: 9.
71. Field A. Cluster analysis. In Discovering Statistics. https://www.discoveringstatistics.com/2017/01/13/cluster-analysis/: 2017.
72. Gaskin J. Two-step Cluster Analysis. In Gaskination. http://youtube.com/Gaskination: 2012.
73. J. G. Validating a two-step cluster analysis - how many clusters? . In. Gaskination's Statistics: www.youtube.com/Gaskination 2015.
74. Gaskin J. Validating K-means cluster analysis in SPSS. In Gaskination. http://youtube.com/Gaskination: 2015.
75. Factor analysis and related techniques. London: Sage Publications Ltd.,1994.